

P-2210

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-285140

(43)Date of publication of application : 13.10.2000

(51)Int.Cl.

G06F 17/30

(21)Application number : 11-343890

(71)Applicant : RICOH CO LTD

(22)Date of filing : 02.12.1999

(72)Inventor : SHIMADA ATSUO
MIYAJI TATSUO
KENMOCHI EIJI
YAMAZAKI MAKOTO
TAKEYA KAZUHISA
NAGATSUKA TETSUO

(30)Priority

Priority number : 10376576

Priority date : 24.12.1998

Priority country : JP

10369589

25.12.1998

11022915

29.01.1999

JP

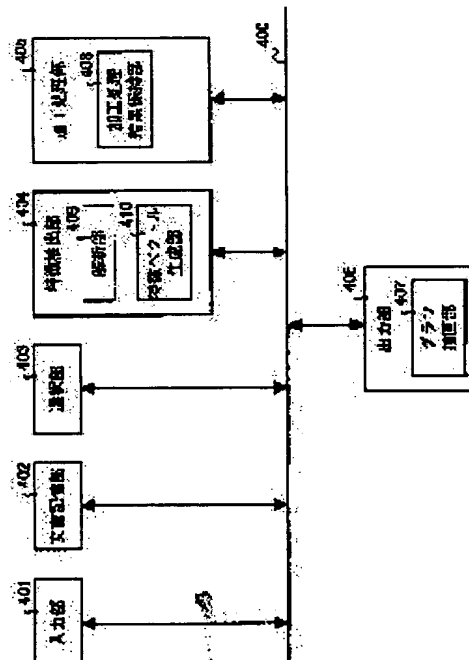
JP

(54) DEVICE AND METHOD FOR PROCESSING DOCUMENT, DEVICE AND METHOD FOR CLASSIFYING DOCUMENT, AND COMPUTER READABLE RECORDING MEDIUM RECORDED WITH PROGRAM FOR ALLOWING COMPUTER TO EXECUTE THESE METHODS

(57)Abstract:

PROBLEM TO BE SOLVED: To make it possible not only to support the result but also to support the whole information analysis in the case of analysis related to the meaning of a document.

SOLUTION: This document processing device is provided with a document storage part 402 for storing inputted document data, a selection part 403 for selecting all the parts or one part of the document data stored in the document storage part 403, a feature extracting part 404 for extracting data on the features of character strings from all the parts or one part of the document data selected by the selection part 403, a work processing part 405 for working all the parts or one part of the document data on the bases of the data on the features of character strings extracted by the feature extracting part 404 and an output part for outputting all the parts or one part of the document data worked by the work processing part 405.



LEGAL STATUS

[Date of request for examination]

(19) 日本国特許庁 (F)

(12) 公 開 特 許 公 報 (A)

(11) 特許出願公開番号

特開2000-285140
(F2000-285140A)

(43) 公開日 平成12年10月13日 (2000. 10. 13)

(61) Int. Cl.⁷ 分類記号
G 0 6 F 17/30 F I
G 0 6 F 15/401 3 1 0 D 5 B 0 7 5
15/40 3 7 0 A
15/401 3 3 0 Z

審査請求 未請求 請求項の数38 O L (全 80 頁)

(21) 出願番号	特開平11-343890	(71) 出願人	00006747 株式会社リコー 東京都大田区中馬込1丁目3番6号
(22) 出願日	平成11年12月2日 (1999. 12. 2)	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(31) 優先権主張番号	特開平10-376576	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(32) 優先日	平成10年12月24日 (1998. 12. 24)	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(33) 優先権主張国	日本 (J P)	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(31) 優先権主張番号	特開平10-368589	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(32) 優先日	平成10年12月25日 (1998. 12. 25)	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(33) 優先権主張国	日本 (J P)	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(31) 優先権主張番号	特開平11-22915	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(32) 優先日	平成11年1月29日 (1999. 1. 29)	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)
(33) 優先権主張国	日本 (J P)	(72) 発明者	嶋田 敦夫 東京都大田区中馬込1丁目3番6号 株式会社リコー内 宮地 達生 東京都大田区中馬込1丁目3番6号 株式会社リコー内 100104190 (74) 代理人 弁護士 西井 昭徳 日本 (J P)

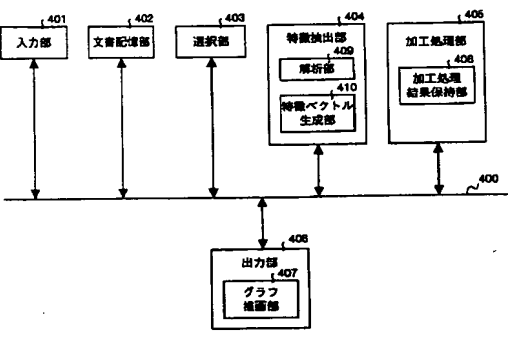
最終頁に続く

(54) 【発明の名称】 文書処理装置、文書分類装置、文書処理方法およびそれらの方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体

(57) 【要約】

【課題】 文 の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことを課題とする。

【解決手段】 入力された文書データを記憶する文書記憶部402と、文 記憶部402により記憶された文書データの全部または一部を選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出部404と、特徴抽出部404により抽出された文字列の特徴に関するデータに基づいて文書データの全部または一部を加工処理する加工処理部405と、加工処理部405により加工処理された文書データの全部または一部を出力する出力部406とを備える。



【特許請求の範囲】

【請求項1】 入力された複数の文書データを所定の形式で表示または印刷するために出力する文書処理装置において、

入力された文書データを記憶する文書記憶手段と、前記文書記憶手段により記憶された文書データの全部または一部を選択する選択手段と、前記選択手段により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、

前記特徴抽出手段により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段と、

前記出力手段により出力された文書データを記憶する文書記憶手段と、前記文書記憶手段により記憶された文書データの全部または一部を選択する選択手段と、前記選択手段により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、

前記特徴抽出手段により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段と、

前記出力手段により出力された文書データを記憶する文書記憶手段と、前記文書記憶手段により記憶された文書データの全部または一部を選択する選択手段と、前記選択手段により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、

前記特徴抽出手段により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段と、

前記出力手段により出力された文書データを記憶する文書記憶手段と、前記文書記憶手段により記憶された文書データの全部または一部を選択する選択手段と、前記選択手段により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、

前記特徴抽出手段により抽出された文字列の特徴に関するデータに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル間の類似度に基づいて文書を分類し、文書の部分集合を生成する分類手段と、

前記分類手段により生成された文書の部分集合の特徴であるクラスラスタ特徴を算出するクラスラスタ特徴算出手段と、前記クラスラスタ特徴算出手段により算出されたクラスラスタ特徴を分類体系の構成要素として記憶する分類体系記憶手段と、

前記分類体系記憶手段により記憶された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、前記ベクトル生成手段により生成された文 特徴ベクトル生成手段と、

た文、特徴ベクトル間の類似度を判断する際の文書表現空間を前記クラスタ選択指示工程により選択された部分集合から算出する特徴量に基づいて修正する文書表現空間修正工程と、

を含み、

前記分類工程は、前記文、表現空間修正工程により修正された文、表現空間を用いて、前記ベクトル修正工程により修正された文、特徴ベクトル間の類似度に基づいて文を分類することを特徴とする請求項32に記載の文分類方法。

【請求項35】 前記分類工程により生成された文書の部分集合に所属する文のすべてあるいは一部が選択された場合に選択されたことを示す選択情報を付与する選択情報付与工程を含み、

前記表示工程は、前記クラスタ特徴を表示するとともに、前記選択情報付与工程により付与された選択情報を表示することを特徴とする請求項31または33に記載の文分類方法。

【請求項36】 前記分類体系生成工程は、前記選択指示工程により選択されたクラスタ特徴のほか、前記文書の部分集合の中から選択された文書の部分集合に所属する文群の全部あるいは一部および/または操作者が作成した情報に基づいて分類体系の構成要素を生成することを特徴とする請求項31～35に記載の文書分類方法。

【請求項37】 文の内部に記しがつて文書群を分類する文分類方法において、
文書データ群を入力し、入力された文書データ群の各文に対して所定の基準に基づき文書の分類をおこない、一つの文データから一つまたは複数の分割文書データを生成し、前記文データと前記分割文書データとの対応を示す文一分割文書対応マップを生成し、前記分割文データを分類し、分割文書分類結果情報と前記文一分割文書対応マップと前記分割文書分類結果情報とを用いて前記文書データの分類結果情報を生成することを特徴とする文分類方法。

【請求項38】 前記請求項24～37のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことを特徴とするコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、入力された複数の文データを所定の形式で表示または印刷するために出力する文処理装置、文書処理方法およびその方法を実行するコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。また、この発明は、入力された複数の文書をその文書の内容に基づいて分類をおこなう、特に文分類の際に算出される分類カテゴリ(体系)を精細化する文分類装置、文分類

方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。

【0002】

【従来の技術】近年、さまざまな文書分類装置や文書検索装置が開発されている。また、インターネット等のネットワーク技術の普及により国内外の大量の電子化文書へのアクセスが可能になり、それに比例して業務上電子的に蓄積される情報の量も飛躍的に拡大した。その中で収集した大量の文書情報を意味あるカテゴリ(体系)に分類する等の知的作業の必要性が高まってきている。

【0003】これらの大量の文書情報を意味的に分類するという作業の目的は、以下のようなものである。まず第1に、検索容易性の向上が考えられる。これは、膨大な文書群を分類する(内容)を手がかりに検索できるため検索が比較的容易になるというものである。

【0004】第2に、情報源全体の把握が考えられる。これは、文書群全体がどのような内容(個々の分類)で構成されているかを把握する。しかし、大量の文書情報を操作者が手動で分類する場合、正確な分類をすることはできず、分類に係る人的、時間的コストが膨大なものになるため、近年の文書の蓄積量の膨大から、文書情報の自動分類装置が提案されるようになってきた。

【0005】文書自動分類装置の従来技術としては、たとえば、特開平7-36897号公報に記載されているように、文書を、単語を特徴とする文書ベクトルとみなし、クラスタリング手法を用いてこれらの文書ベクトルを群分けし、群分けした文書ベクトルに基づいて文書の自動分類をおこなうものがある。

【0006】また、Projections for Efficient Document Clustering (著者: Hinrich Schutze and Graing Silverstein, 学名: ACM, 論文名: Proceedings of SIGIR, ページ: 74-81, 発行年: 1997) 1)においては、潜在的意味空間において文書

分類を実施しているものがある。そのほかの方法としては、種々のプロローチを用いる方法が考えられる。【0007】また近年、インターネットなどの普及により、大量の文書群へのアクセスが可能になり、その結果、その文書群をさまざまな利用者の意図に基づいて、かつ、効率的に利用できるようにする必要は高まっている。そのため、大量の文書群を意味のあるカテゴリに分類し、文書群の構造を把握するという知的作業がおこなわれ始めている。しかし、このような分類作業を人手によりおこなう場合、その人的および時間的なコストが膨大なものになり、また、分類のための知識を分類者のみが有することになるため、分類担当者が代わると分類基準も変わってしまうことになる。

【0008】そのため、文群を人間が分類するような

分類基準で自動的に分類しうる文書分類装置が望まれており、文書分類装置としては、たとえば、特開平7-14572号公報に記載されているように、文書から自動的に単語の特徴ベクトルを抽出し、その特徴ベクトルをもとに文書分類することで、意味的な真なりを用いた自動分類を可能にするものがある。

【0009】

【発明が解決しようとする課題】しかしながら、上記従来技術の文書分類装置は、本質的には単語で構成される多次元空間に布置した文書を統計的な分類をする方法であるため、分類結果は単語のいわゆる重なりという観点から統計的に求められたものにすぎず、分類の結果、算出される各クラスタ(分類された個々の文書の部分集合)が操作者(利用者)に理解不能な場合がある。

【0010】また、どのような分類結果が最適かは、分類対象の文書集合の特徴や、利用者の作業の目的に依存するため、最適な分類結果について定数することが困難であるという問題点があった。特に、上記情報群全体の把握に関し、多様な操作者の意図により要求される分類も異なるため、一度の分類作業で、操作者の所望する結果を得ることが困難であるという問題点があった。

【0011】このように、文書分類の結果は、多くのいわゆるノイズを含んだものであると解釈することができ、その一部にすぎないのが操作者にとって有益な場合が多いという問題点があった。

【0012】また、これらの従来技術においては、文書の構成単位を考慮していないため、区画が一つまたは複数の段落記号やタイトルなどにより区切られる構造を持つ場合には、一つの文書の中に複数の話題や意味が含まれてしまい、その結果、利用者がその分類カテゴリを理解し難くなったり、また、ある特定の話題や特定の意味に限定されたカテゴリになったり、利用者の意図するカテゴリとは異なるカテゴリに分類されてしまうという問題が生じている。

【0013】なお、特開平6-176064号公報に示された文脈依存自動分類装置には、文書の段落構造を考慮した文書自動分類をおこなうことにより分類精度を高めようとするものが開示されているが、本質的に上記の問題を解決するものではない。

【0014】また、上記従来技術の文書分類装置や文書検索装置等の文書処理装置は、単に文書を分類する、あるいは文書を検索する機能を有するのみで、その結果を用いてさらなる分析をおこない、文書群に内在する隠れた情報の解析をおこなうことについては何ら考慮がされておらず、文書群に内在する隠れた情報の解析は別の解析装置を用いておこなわなければならないという問題点があった。

【0015】また、情報分析をおこなう操作者が分類作業や検索作業をおこなうのは、これらの作業において、結果は目的ではなく、単に情報分析作業の途中経過

にすぎないからである。通常は、その後、さらに結果を把握しやすくするために、元の文書に含まれる情報を最大限に活用し、結果の並び替えをおこなったり、集計・統計処理を施したり、結果をもとに表の形式にまとめたリ、さらにはグラフ化したりというようなさまざまな処理を繰り返しておこない、意味ある情報分析結果を導き出す必要がある。

【0016】また、数値データを対象とする情報の分析作業において、数計算ソフトウェアが用いられる場合があるが、数計算ソフトウェアは、元来、数値データの取扱いを意図して開発されたものであり、文字データ、特に文書の意味に係るような分析作業においては十分な効果を発揮することはできなかった。

【0017】この発明は、上述した従来例による問題点を解消するため、文書の意味に係るような分析作業において、単に分類作業や検索作業などを限定された機能としておこない、その結果を出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる文処理装置、文書処理方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを第1の目的とする。

【0018】またこの発明は、上述した従来例による問題点を解消するため、任意の文、集合にどのような内容が含まれるかを漸次的に収集することができる文分類装置、文書分類方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを第2の目的とする。

【0019】またこの発明は、上述した従来例による問題点を解決するため、一つの文の中に複数の話題や意味が含まれている場合に、ある特定の話題や意味に限定されたカテゴリに分類されたり、利用者の意図するカテゴリとは異なるカテゴリに分類されることがないことにより、利用者かその分類カテゴリをよく理解できる文分類装置、文書分類方法およびその方法をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体を提供することを第3の目的とする。

【0020】

【課題を解決するための手段】上述した課題を解決し、目的を達成するため、請求項1の発明に係る文処理装置は、入力された複数の文データを所定の形式で表示または印刷するために出力する文処理装置において、入力された文書データを記憶する文記憶手段と、前記文書記憶手段により記憶された文データの全部または一部を選択する選択手段と、前記選択手段により選択された文データの全部または一部から文字列の特徴に関するデータと、前記文書データの文字列の特徴に関するデータとに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文データの全部または一部を出力する出力手段と、そ

備えたことを特徴とする。

【0021】この請求項1の発明によれば、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0022】また、請求項2の発明に係る文書処理装置は、請求項1の発明において、前記出力手段が、前記加工処理手段により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定手段と、前記項目値設定手段により設定された項目値ごとに前記文データの全部または一部を集計する集計手段と、を備え、前記文書データの全部または一部を、項目値を少なくとも一つの軸とする表形式に展開して出力することを特徴とする。

【0023】この請求項3の発明によれば、簡易な操作で加工処理の結果をクロス表として表すことができ、情報の内容の把握を容易におこなうことができ、さらに、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0024】また、請求項3の発明に係る文書処理装置は、請求項1または2の発明において、前記出力手段が、さらに、前記加工処理手段により加工処理された文データの全部または一部を、前記加工処理手段により加工処理される前の文データの全部または一部とともに出力することを特徴とする。

【0025】この請求項3の発明によれば、加工処理すべき対象データとその他のデータが同時に表示され、それを確認することにより、加工処理の対象範囲の決定を正確かつ容易におこなうことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0026】また、請求項4の発明に係る文書処理装置は、請求項1～3の発明において、前記文書記憶手段が、さらに、前記加工処理手段により加工処理された文データの全部または一部を記憶することを特徴とする。

【0027】この請求項4の発明によれば、以後、他のデータと同様に扱うことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0028】また、請求項5の発明に係る文書処理装置は、請求項1～4の発明において、前記選択手段が、さらに、前記出力手段により出力された文書データの全部または一部を選択することを特徴とする。

【0029】この請求項5の発明によれば、出力手段により出力された文データの全部または一部をさらなる分析の対象とすることができ、多形で高度な情報分析作

業ができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0030】また、請求項6の発明に係る文書処理装置は、請求項1～5の発明において、前記文書記憶手段が、さらに、前記加工処理の内容に関するデータを記憶することを特徴とする。

【0031】この請求項6の発明によれば、加工処理の内容に関するデータの損失を防止し、当該データの管理が容易になるだけでなく、加工処理に用いた設定とそれによる処理結果を関連付けて把握することができ、さらに、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0032】また、請求項7の発明に係る文書分類装置は、文書の内容に基づいて文書の分類をおこなう文書分類装置において、文書データをを入力する入力手段と、前記入力手段により入力された文書データを解析して言語解析情報を得る言語解析手段と、前記言語解析手段により得られた言語解析情報に基づいて前記文書データに対する文書特徴ベクトルを生成するベクトル生成手段と、前記ベクトル生成手段により生成された文書特徴ベクトルの類似度に基づいて文書を分類し、文書の部分集合を生成する分類手段と、前記分類手段により生成された文書の部分集合の特徴であるクラスラスタ特徴を算出するクラスラスタ特徴算出手段と、前記クラスラスタ特徴算出手段により算出されたクラスラスタ特徴を分類体系の構成要素として記憶する分類体系記憶手段と、を備えたことを特徴とする。

【0033】この請求項7の発明によれば、クラスラスタを得ることができるとともに、クラスラスタ間の類似度等を用いて、クラスラスタの内容に基づくクラスラスタの構造化・体系化をおこなうことができる。

【0034】また、請求項8の発明に係る文書分類装置は、文書の内容に基づいて文書の分類をおこなう文書分類装置において、文書データを入力する入力手段と、前記入力手段により入力された文書データを解析して言語解析情報を得る言語解析手段と、前記言語解析手段により得られた言語解析情報に基づいて前記文書データに対する文書特徴ベクトルを生成するベクトル生成手段と、前記ベクトル生成手段により生成された文書特徴ベクトル間の類似度に基づいて文書を分類し、文書の部分集合を生成する分類手段と、前記分類手段により生成された文書の部分集合の特徴であるクラスラスタ特徴を算出するクラスラスタ特徴算出手段と、前記クラスラスタ特徴算出手段により算出されたクラスラスタ特徴を表示する表示手段と、前記分類手段により生成された文書の部分集合の中から所望の部分集合を選択するクラスラスタ選択指示手段と、前記クラスラスタ選択指示手段により選択された文の部分集合を

分類体系の構成要素として記憶する分類体系記憶手段と、を備えたことを特徴とする。

【0035】この請求項8の発明によれば、選択されたクラスラスタのみを用いており、より操作者の意図したものに近いクラスラスタの構造化・体系化をおこなうことができる。

【0036】また、請求項9の発明に係る文書分類装置は、請求項8の発明において、前記ベクトル生成手段により生成された文書特徴ベクトルを記憶する文書特徴ベクトル記憶手段と、前記文書特徴ベクトル記憶手段により記憶された文書特徴ベクトルを、前記クラスラスタ選択指示手段により選択された部分集合に属する文書の文書特徴ベクトルを除去したのりとなるように修正するベクトル修正手段と、を備え、前記分類手段が、前記ベクトル修正手段により修正された文書特徴ベクトルに基づいて文書を分類することを特徴とする。

【0037】この請求項9の発明によれば、既知になったクラスラスタの影響を排除した新たなクラスラスタを生成することができ、

【0038】また、請求項10の発明に係る文書分類装置は、請求項8の発明において、前記ベクトル生成手段により生成された文書特徴ベクトルを記憶する文書特徴ベクトル記憶手段と、前記文書特徴ベクトル記憶手段により記憶された文書特徴ベクトル間の類似度を判断する際の文書表現空間を前記クラスラスタ選択指示手段により選択された部分集合から算出する特徴量に基づいて修正する文書表現空間修正手段と、を備え、前記分類手段が、前記文書表現空間修正手段により修正された文書表現空間を用いて、前記ベクトル生成手段により生成された文書特徴ベクトル間の類似度に基づいて文書を分類することを特徴とする。

【0039】この請求項10の発明によれば、前回の分類実行の結果、操作者に選択されたクラスラスタの形成特徴を次の回の分類実行時に排除することができ、排除した状態で新たなクラスラスタを生成することができ、

【0040】また、請求項11の発明に係る文書分類装置は、請求項9の発明において、前記ベクトル生成手段により生成された文書特徴ベクトルを記憶する文書特徴ベクトル記憶手段と、前記文書特徴ベクトル記憶手段により記憶された文書特徴ベクトル間の類似度を判断する際の文書表現空間を前記クラスラスタ選択指示手段により選択された部分集合から算出する特徴量に基づいて修正する文書表現空間修正手段と、を備え、前記分類手段が、前記文書表現空間修正手段により修正された文書表現空間を用いて、前記ベクトル修正手段により修正された文書特徴ベクトル間の類似度に基づいて文書を分類することを特徴とする。

【0041】この請求項11の発明によれば、既知になったクラスラスタの影響を排除し、かつ、前回の分類実行の結果、操作者に選択されたクラスラスタの形成特徴を次の回の

分類実行時に排除することができ、排除した状態で新たなクラスラスタを生成することができ、

【0042】また、請求項12の発明に係る文分類装置は、請求項8または10の発明において、前記分類手段により生成された文の部分集合に所属する文のすべてあるいは一部が選択された場合に選択されたことを示す選択情報を付与する選択情報付与手段とを備え、前記表示手段が、前記クラスラスタ特徴を表示するとともに、前記選択情報付与手段により付与された選択情報を表示することを特徴とする。

【0043】この請求項12の発明によれば、多量に利用される文書の識別性および一度も選択されない文書の識別性を向上させることができる。

【0044】また、請求項13の発明に係る文分類装置は、請求項8～12の発明において、前記分類体系記憶手段が、前記選択指示手段により選択された文の部分集合に属する全部あるいは一部の文書のほか、クラスラスタ特徴および/または操作者が作成した任意の情報进行分类体系の構成要素として記憶することを特徴とする。

【0045】この請求項13の発明によれば、クラスラスタの内容把握を容易にし、かつ、操作者独自の分類体系を簡易に生成できるので、分類体系の利用価値を向上させることができる。

【0046】また、請求項14の発明に係る文書分類装置は、文書の内容にしたがって文書を分類する文分類装置において、文書データを群を入力する文書入力手段と、入力された文書データ群のなかの、一つの文データから一つまたは複数の分割文データを生成する文分割手段と、前記文データと前記分割文データとの対応を示す文書一分割文書対応マッピングを生成する文一分割文書対応マッピング生成手段と、前記分割文書データによる分割文書分類手段と、前記分割文書分類手段による分割結果に基づいて分割文書分類結果情報とを生成する分割文書分類結果生成手段と、前記文一分割文書対応マッピングと前記分割文書分類結果情報とを用いて前記文データの分類結果情報とを生成する文分類結果生成手段と、を備えたことを特徴とする。

【0047】この請求項14の発明によれば、一つの文書の中に複数の話題や意味が含まれている場合に、ある特定の話題や意味に限定されたカテゴリに分類されたり、利用者の意図のカテゴリとは異なるカテゴリに分類されたりすることがなく、したがって、利用者自身が分類カテゴリをよく理解でき、また、分割前文（所属文書）中の分割文書の位置なども示されるので、利用者は文書群中の読みたい部分を効率的に読むことができる。

【0048】また、請求項15の発明に係る文分類装置は、請求項14の発明において、前記文データを保存する文保存手段と、前記分割文データを

分割文、保存手段と、前記文書一分割文書対応ウェブ生成手段により生成された文書一分割文書対応ウェブを保存する文、一分割文 対応ウェブ保存手段と、を備えたことを特徴とする。

【0049】この請求項15の発明によれば、分割文書データおよび文、一分割文書対応ウェブを再生成することとに、同一の文、データに対して、分類、分類方法、または分類時の諸設定などパラメータの異なる分類結果を動的に求めることができる。また、文書データを分類し、分類結果を生成するために必要なデータが保存されることにより、利用者が分類作業に対して時間的な自由度を持つことができるし、過去に行った文書分類の再分析を任意の時間間におこなうこともできる。

【0050】また、請求項16の発明に係る文書分類装置は、請求項15の発明において、前記分割文書分類結果生成手段により生成された分割文書分類結果情報を保存する分割文、分類結果保存手段を備えたことを特徴とする。

【0051】この請求項16の発明によれば、請求項15の発明の効果に加え、一度分類を実行すれば、その分類結果をテキスト表現や数表現やグラフ表現などさまざまな形式で取得することができる。また、分割文書分類結果情報が保存されることにより、分類の実行作業および分類結果の分析作業において、分類の実行作業および分類結果の分析作業において、利用者が時間的な自由度を持つことができるし、過去に行った文書分類結果の再分析をさまざまな表現形式で任意の時間におこなうこともできる。

【0052】また、請求項17の発明に係る文書分類装置は、請求項14～16の発明において、前記文書分割手段により生成される複数の分割文書データには分割前の文、データそのものを含むことを特徴とする。

【0053】この請求項17の発明によれば、利用者は、分割された文、データを分類することを得られる詳細な文、データの分類構造だけでなく、分割前の文、データ自体を分類した結果として得られる概略的でクロな分類構造の融合した分類構造を得ることができる。

【0054】また、請求項18の発明に係る文書分類装置は、請求項14～17の発明において、前記文書分割手段が、文、データの構造情報を基に文書データを分割する構成にしたことを特徴とする。

【0055】この請求項18の発明によれば、異なった疑問の分類等を通じておこなうことができ、したがって、文、データの詳細な分類構造がわかる文書分類を通じておこなうことができる。

【0056】また、請求項19の発明に係る文書分類装置は、請求項14～17の発明において、前記文書データを含まれる要素を抽出する要素抽出手段と、前記文、要素抽出手段により抽出された要素に付随する要素付随情報を抽出する要素付随情報抽出手段と、を備え、

前記文書分割手段が、前記文書要素抽出手段により抽出された要素、または前記要素と前記要素付随情報抽出手段により抽出された要素付随情報とを用いて前記文書データを分割する構成にしたことを特徴とする。

【0057】この請求項19の発明によれば、文書データの詳細な分類構造がわかる文書分類を通じておこなうことができる。

【0058】また、請求項20の発明に係る文書分類装置は、請求項14～17の発明において、前記文書分割手段が、指示された指定範囲にしたがって文書データの分割をおこなう構成にしたことを特徴とする。

【0059】この請求項20の発明によれば、利用者の意図に合い、かつ文書データの詳細な分類構造がわかる文書分類をおこなうことができる。

【0060】また、請求項21の発明に係る文書分類装置は、請求項14～17において、前記文書分割手段が、文書データ中の文字数、文数、または文字数と文数の両方を基に文書データを分割する構成にしたことを特徴とする。

【0061】この請求項21の発明によれば、疑問の異なった内容などが異なった文書として分類される可能性が高くなり、したがって、この発明でも文書データの詳細な分類構造がわかる文書分類をおこなうことができる。

【0062】また、請求項22の発明に係る文書分類装置は、請求項14～21の発明において、前記文書分類結果生成手段が、文書データを示す情報および前記文書データに付随する代表的情報を、分類結果情報として抽出して提示する構成にしたことを特徴とする。

【0063】この請求項22の発明によれば、利用者は文書データの詳細な分類構造の概要や全体的な構造を容易に把握することができる。

【0064】また、請求項23の発明に係る文書分類装置は、請求項22の発明において、前記文書分類結果生成手段が、分割文書データを示す情報および前記分割文書データに付随する代表的情報を、分類結果情報として、抽出して提示する構成にしたことを特徴とする。

【0065】この請求項23の発明によれば、利用者は文書データの詳細な分類構造の概要や全体的な構造とともにどの分割文書が範囲に当該カテゴリに分類されたかというようなことも容易にわかる。

【0066】また、請求項24の発明に係る文書処理方法は、入力された複数の文書データを所定の形式で表示または印刷するために出力する文書処理方法において、入力された文書データを記憶する文書記憶工程と、前記文書記憶工程とより記憶された文書データの全部または一部を選択する選択工程と、前記選択工程により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出工程と、前記特徴抽出工程により抽出された文字列の特徴に関するデータに基づ

いて前記文書データの全部または一部を加工処理する加工処理工程と、前記加工処理工程により加工処理された文書データの全部または一部を出力する出力工程と、を含んだことを特徴とする。

【0067】この請求項24の発明によれば、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0068】また、請求項25の発明に係る文書処理方法は、請求項24の発明において、前記出力工程が、前記加工処理工程により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定工程と、前記項目値設定工程により設定された項目値ごとに前記文書データの全部または一部を集計する集計工程と、を含み、前記文書データの全部または一部を、項目値を少なくとも一つの軸とする表形式に展開して出力することを特徴とする。

【0069】この請求項25の発明によれば、簡易な操作で加工処理の結果をクロス表として表示ことができ、情報の内容の把握を容易におこなうことができることから、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0070】また、請求項26の発明に係る文書処理方法は、請求項24または25の発明において、前記出力工程が、さらに、前記加工処理工程により加工処理された文書データの全部または一部を、前記加工処理工程により加工処理される前の文書データの全部または一部とともに出力することを特徴とする。

【0071】この請求項26の発明によれば、加工処理すべき対象データとその他のデータが同時に表示され、それを確認することにより、加工処理の対象範囲の決定を正確かつ容易におこなうことができることから、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0072】また、請求項27の発明に係る文書処理方法は、請求項24～26の発明において、前記文書記憶工程が、さらに、前記加工処理工程により加工処理された文書データの全部または一部を記憶することを特徴とする。

【0073】この請求項27の発明によれば、以後、他のデータと同様に扱うことができることから、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0074】また、請求項28の発明に係る文書処理方法は、請求項24～27の発明において、前記選択工程が、さらに、前記出力工程により出力された文、データの全部または一部を選択することを特徴とする。

【0075】この請求項28の発明によれば、出力手段により出力された文書データの全部または一部をさらなる分析の対象とすることができ、多岐で高度な情報分析作業ができることから、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0076】また、請求項29の発明に係る文、処理方法は、請求項24～28の発明において、前記文、記憶工程が、さらに、前記加工処理の内容に関するデータを記憶することを特徴とする。

【0077】この請求項29の発明によれば、加工処理の内容に関するデータの紛失を防止し、当該データの管理が容易になるだけでなく、加工処理に用いた設定とそれによる処理結果を関連づけて把握することができることから、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことができる。

【0078】また、請求項30の発明に係る文、分類方法は、文書の内容に基づいて文、の分類をおこなう文、分類方法において、文、データを入力する入力工程と、前記入力工程により入力された文、データを解析して言語解析情報を得る言語解析工程と、前記言語解析工程により得られた言語解析情報に基づいて前記文、データに対する文書特徴ベクトルを生成するベクトル生成工程と、前記ベクトル生成工程により生成された文、特徴ベクトル間の類似度に基づいて文、を分類し、文、の部分集合を生成する分類工程と、前記分類工程により生成された文書の部分集合の特徴であるクラスラスタ特徴を算出するクラスラスタ特徴算出工程と、前記クラスラスタ特徴算出工程により算出されたクラスラスタ特徴に基づいて分類体系の構成要素を生成する分類体系生成工程と、を含んだことを特徴とする。

【0079】この請求項30の発明によれば、クラスラスタを得ることができるとともに、クラスラスタ間の類似度等を用いて、クラスラスタの内容に基づくクラスラスタの構造化・体系化をおこなうことができる。

【0080】また、請求項31の発明に係る文、分類方法は、文書の内容に基づいて文、の分類をおこなう文、分類方法において、文、データを入力する入力工程と、前記入力工程により入力された文書データを解析して言語解析情報を得る言語解析工程と、前記言語解析工程により得られた言語解析情報に基づいて前記文、データに対する文書特徴ベクトルを生成するベクトル生成工程と、前記ベクトル生成工程により生成された文書特徴ベクトル間の類似度に基づいて文、を分類し、文、の部分集合を生成する分類工程と、前記分類工程により生成された文書の部分集合の特徴であるクラスラスタ特徴を算出するクラスラスタ特徴算出工程と、前記クラスラスタ特徴算出工程により算出されたクラスラスタ特徴を表示する表示工程と、

前記分類工程により生成された文書の部分集合の中から所望の部分集合を選択するク拉斯タ選択指示工程と、前記ク拉斯タ選択指示工程により選択されたク拉斯タ特徴に基づいて分類体系の構成要素を生成する分類体系生成工程と、をきんだことを特徴とする。

【0081】この請求項31の発明によれば、選択されたクラスタのみを用いており、より操作者の意図したものに近いクラスタの構造化、体系化をおこなうことができる。

【0082】また、請求項32の発明に係る文書分類方法では、請求項31の発明において、前記ベクトル生成工程により生成された文 特徴ベクトルを、前記ク拉斯タ選択指示工程により選択された部分集合に属する文書の文 特徴ベクトルを除去したのこりとなるように修正するベクトル修正工程と、を含み、前記分類工程が、前記ベクトル修正工程により修正された文書特徴ベクトルに基づいて文 を分類することを特徴とする。

【0083】この請求項32の発明によれば、既知になったクラスタの影響を排除した新たなクラスタを生成することができ、

【0084】また、請求項33の発明に係る文書分類方法では、請求項31の発明において、前記ベクトル生成工程により生成された文 特徴ベクトル間の類似度を判断する際、文 表現空間を前記ク拉斯タ選択指示工程により選択された部分集合から算出する特徴量に基づいて修正する文 表現空間修正工程と、を含み、前記分類工程が、前記文 表現空間修正工程により修正された文書表現空間を用いて、前記ベクトル生成手段工程により生成された文 特徴ベクトル間の類似度に基づいて文書を分類することを特徴とする。

【0085】この請求項33の発明によれば、前回の分類実行の結果、操作者に選択されたクラスタの形成特徴を次の分類実行時に排除することができ、排除した状態で新たなクラスタを生成することができる。

【0086】また、請求項34の発明に係る文書分類方法では、請求項32の発明において、前記ベクトル生成工程により生成された文 特徴ベクトル間の類似度を判断する際、文 表現空間を前記ク拉斯タ選択指示工程により選択された部分集合から算出する特徴量に基づいて修正する文 表現空間修正工程と、を含み、前記分類工程が、前記文 表現空間修正工程により修正された文書表現空間を用いて、前記ベクトル修正工程により修正された文 特徴ベクトル間の類似度に基づいて文書を分類することを特徴とする。

【0087】この請求項34の発明によれば、既知になったクラスタの影響を排除し、かつ、前回の分類実行の結果、操作者に選択されたクラスタの形成特徴を次の分類実行時に排除することができ、排除した状態で新たなクラスタを生成することができる。

【0088】また、請求項35の発明に係る文 分類方

法は、請求項31または33の発明において、前記分類工程により生成された文書の部分集合に所属する文書のすべてあるいは一部が選択された場合に選択されたことを示す選択情報を付与する選択情報付与工程を含み、前記表示工程が、前記ク拉斯タ特徴を表示するとともに、前記選択情報付与工程により付与された選択情報を表示することを特徴とする。

【0089】この請求項35の発明によれば、多量に利用される文書の識別性および一度も選択されない文書の識別性を向上させることができる。

【0090】また、請求項36の発明に係る文書分類方法では、請求項31～35の発明において、前記分類体系生成工程が、前記選択指示工程により選択されたク拉斯タ特徴のほか、前記文書の部分集合の中から選択された文書の部分集合に所属する文書群の全部あるいは一部および／または操作者が作成した情報に基づいて分類体系の構成要素を生成することを特徴とする。

【0091】この請求項36の発明によれば、クラスタの内容把握を容易にし、かつ、操作者独自の分類体系を簡単に生成できるので、分類体系の利用価値を向上させることができる。

【0092】また、請求項37の発明に係る文書分類方法では、文書の内容に基づいた文書群を分類する文書分類方法において、文書データ群を入力し、入力された文書データ群の各文書に対して所定の基準に基づき文書の分割をおこない、一つの文書データから一つまたは複数の分割文書データを生成し、前記文書データと前記分割文書データとの対応を示す文書一分割文書対応マップを生成し、前記分割文書データを分類し、分割文書分類結果情報と、前記文書一分割文書対応マップと前記分割文書分類結果情報を用いて前記文書データの分類結果情報を生成することを特徴とする。

【0093】この請求項37の発明によれば、一つの文書の中に複数の話題や意味が含まれている場合、ある特定の話題や意味に関連されたカテゴリに分類され、利用者の意図するカテゴリとは異なるカテゴリに分類されたりすることがなく、したがって、利用者がその分類カテゴリをより理解できる。また、分割前文書（所属文書）中の分割データの位置なども示されるので、利用者は文書群中の読みたい部分を効率的に読むことができる。

【0094】また、請求項38の発明に係る記憶媒体は、請求項24～31に記載された方法をコンピュータに実行させるプログラムを記録したことで、そのプログラムを機械読み取り可能になり、これによって、請求項24～37の動作をコンピュータによって実現することが可能である。

【0095】

【発明の実施の形態】以下に添付図面を参照して、この発明に係る文 処理装置、文 処理方法およびその方法

をコンピュータに実行させるプログラムを記録したコンピュータ読み取り可能な記録媒体の好適な実施の形態を詳細に説明する。

【0096】【実施の形態1】まず、この発明の実施の形態1による文書処理装置を構成する情報処理システム全体のハードウェア構成を説明する。図1は、実施の形態1による文書処理装置を構成する情報処理システム全体のハードウェア構成を示す説明図である。

【0097】図1において、実施の形態1による文書処理装置を構成する情報処理システムは、サーバ／クライアント方式で構成されている。すなわち、サーバ101と複数のクライアント102がネットワーク103によって接続されている。クライアント102は、分類データ等の加工データの生成、サーバ101への指示、分類結果等の加工処理結果の表示などをおこなう。一方、クライアント102からの指示にしたがって、サーバ101は文書（テキスト）分類等の加工処理を膨大な数値演算によりおこない、その処理の結果をクライアント102へ送る。

【0098】分類処理の場合、より具体的には、サーバ101においては、テキスト分類処理（前処理、ク拉斯タリング処理）がおこなわれ、クライアント102においては、分類データ生成、処理実行指示、テキスト分類結果表示等がおこなわれる。サーバ101における処理は、上流のように、「前処理」と「分類処理」の二つに分かれており、その処理はデータによっては非常に負荷が大きくなる。したがって、サーバ101は「前処理」と「分類処理」がそれぞれ一つずつしか処理をおこなわないようにクライアント102が処理受付リストを作成して管理する。

【0099】また、サーバ101とクライアント102との間のデータのやりとりはファイル共有という方法を用いる。すなわち、分類処理等の加工処理に用いるファイルはサーバ101上の共有フォルダに作成することにより両者はデータのやりとりをおこなう。したがって、クライアント102からはサーバ101の共有フォルダをネットワーク共有して利用することが可能である。

【0100】つぎに、サーバ101およびクライアント102のハードウェア構成について説明する。図2は、実施の形態1による文書処理装置を構成する情報処理システムにおけるサーバ101のハードウェア構成を示す説明図である。サーバ101は、たとえばワーキングステーション（WS）等が用いられる。【0101】図2において、201はサーバ101全体を制御するCPUを、202はアプリケーション等を記憶したROMを、203はCPU201のワーキングエリアとして使用されるRAM203を、204は通信回線205を介してネットワーク103に接続され、そのネットワーク103と内部のインタフェースを司るイン

ターフェイス（I/F）を、206はデータを記憶するデータの装置を示している。200は上記各部を結合させるためのバスを示している。

【0102】そのほか、文書情報、画像情報、機能情報等を表示するディスプレイ208や、データを入力するためのキーボード209およびマウス210等が同様に接続されている。さらに、ディスプレイ装置206には、クライアント102との間のデータのやりとりをするための共有フォルダ207が設けられている。

【0103】また、図3は、実施の形態1による文 処理装置を構成する情報処理システムにおけるクライアント102のハードウェア構成を示す説明図である。クライアント102は、たとえばパーソナルコンピュータ（PC）等が用いられる。

【0104】図3において、301はシステム全体を制御するCPUを、302はアプリケーション等を記憶したROMを、303はCPU301のワーキングエリアとして使用されるRAMを、304はCPU301の制御にしたがってHD（ハードディスク）305に対するデータのリード／ライトを制御するHDD（ハードディスクドライブ）を、306はHDD304の制御で書き込まれたデータを記憶するHDDを、306はCPU301の制御にしたがってFD（フロッピーディスク）307に対するデータのリード／ライトを制御するFD（フロッピーディスクドライブ）を、307はFD306の制御で書き込まれたデータを記憶する着脱自在のFDを、308はドキュメント、画像、機能情報等を表示するディスプレイをそれぞれ示している。

【0105】また、309は通信回線310を介してネットワーク103に接続され、そのネットワーク103と内部のインタフェースを司るインタフェース（I/F）を、311は文字、数値、各種指示等の入力のためのキーを備えたキーボードを、312はカールの移動や範囲選択、あるいは表示画面に示されたアイコンやボタンの押下やインポートの移動やサイズの変更等をおこなうマウスを、313はOCR（Optical Character Reader）機能を備えた画像を光学的に読み取るスキャナを、314は分類結果を含むデータの内容等を印刷するプリンタを、315は上記各部を結合するためのバスをそれぞれ示している。また、HD305にはワーキング領域のアリケーション316が記憶されている。

【0106】つぎに、実施の形態1による文 処理装置の機能構成について説明する。図4は、実施の形態1による文書処理装置の構成を機能的に示すブロック図である。図4において、文 処理装置は、入力部401と、文書記憶部402と、選択部403と、特徴抽出部404と、加工処理部405と、出力部406を含む構成である。

【0107】入力部401、文 記憶部402、選択部

【0164】また、加工処理部405への入力データとして、特徴抽出部404から出力されたデータだけではなく、選択部403により選択されたデータも含めるとが得意な。これにより、文字列の特徴抽出を必要としないデータや、加工処理結果の数値に對してもさらなる加工処理を施すことができるので、より多彩で高度な情報分析が可能となる。

【0165】図24～図24は、実施の形態1による文書処理装置の出力部406による画面表示の別の例を示す説明図である。図24において、「番号」、「受付日」、「番案所」、「車種」、「年式」、「内容」の他に、分類処理により得られた結果である「クラス番号」2101が表示されている。

【0166】さらに、図24においては、選択部403により「クラス番号」2101が選択されており、「クラス番号」2101に関するデータが反転表示されている。選択された「クラス番号」2101をキーとして、加工処理部405により並べ替え処理をおこなうよう指示をする。

【0167】並べ替え処理の指示により、並べ替え処理がおこなわれた結果を表示しているのが図22である。図22においては、「クラス番号」が「1」のものが集まって表示されるように並べ替えられ、それに続き、「クラス番号」が「2」のものが集まって表示されるように並び替えられる。

【0168】具体的に、図22においては、「クラス番号」が「1」である「番号」が「12」、「11」、「15」、「23」、「35」、「54」、「63」、「73」、「83」の順で並べ替えられ、それに続き「クラス番号」が「2」である「番号」が「14」、「18」、「21」、「27」、「37」、... が表示されていることがわかる。

【0169】つぎに、項目「車種」の順で、「クラス番号」が「1」に属するものを選択する。図23においては、項目「車種」の順で、「クラス番号」が「1」に属するものが選択され、その選択領域2301が反転表示されていることを示している。このように、すでに「クラス番号」により並べ替えがおこなわれており、同一クラスに属するものが集まって表示されているので、画面上的連続した領域として容易に選択することができる。

【0170】つぎに、選択領域2301について車種別の発生頻度の棒グラフを表示させたのが、図24である。図24において、棒グラフ表示領域2401には、選択領域2301によって選択された「クラス番号」が「1」である9つの文書が選択され、その9つの文書を車種別に棒グラフ化したものが表示される。

【0171】このように、加工処理の対象を柔軟かつ容易に選択でき、選択された対象について多様な加工処理をおこなうことができ、また、その加工処理結果も次回

の加工処理の対象とすることができるので、高度な情報分析作業が可能となる。

【0172】このように、分類等の文字列の特徴量を抽出して、その特徴量を用いておこなう加工処理を実施した後に多様な加工処理をおこなう例を示したが、事前に多様な処理をおこなうことができるようにしてもよい。

【0173】たとえば、「車種」の項目を選択し、これをキーとして並べ替えをおこなった後、集まったある車種、たとえば、「ABC1600」に對して分類処理をおこなうこともできる。また、入力部401により入力された文書が保管等の処理を含んでいる場合、分類等の文字列の特徴量を抽出して、その特徴量を用いて加工処理をおこなう前に、たとえば、文字列の検索・置換処理をおこなって、誤字を一括して修正し、より好適な結果が得られるようにデータを整えることもできる。

【0174】図25は、実施の形態1による文書処理装置の文書記憶部402の詳細な構成を示すブロック図である。図25において、文書記憶部402は、設定値記憶部2501および設定値送受信部2502を含んでいる。設定値記憶部2501には、文書を分類する際の分類数等の分類情報記憶部2503をはじめとするさまざまな設定値、すなわち文書処理装置の動作に必要な設定値に関する情報を記憶する記憶部を備えている。これにより設定値に関する情報は、文書情報とともに記憶することができるとなる。

【0175】また、設定値送受信部2502は、設定値記憶部2501によって記憶された設定値に関する情報を他の情報処理装置へ送信する。また、設定値送受信部2502は、他の情報処理装置からの設定値に関する情報を受信する。設定値送受信部2502により受信された設定値に関する情報は、設定値記憶部2501によって記憶される。

【0176】記憶された設定値に関する情報は、後に文書を再度読み込んだときに同時に読み込まれた設定値記憶部2501に記憶される。この設定値に関する情報は操作者が所定の操作をすることにより参照することができる。これにより、設定値に関する情報を文書とともに保存・管理することが可能となるので、設定値に関する情報の紛失を防ぎ、好適な設定値を数回再利用することができるとなる。

【0177】図26～図28は、実施の形態1による文書処理装置の出力部406による画面表示の別の例を示す説明図である。図26において、まず、操作者が分類をおこなうべき対象である「内容」を表示画面上で選択する。それにより選択領域2601が反転表示される。つぎに、メニュー・バー2603から、分類処理ボタン2603を選択すると、分類処理に必要な分類数、すなわち、対象をいくつに分類するかについての問い合わせ画面2604が表示される。

【0178】操作者が問い合わせ画面2604において分類数を入力すると、この分類数に関する情報が文書記憶部402に記憶される。図26においては、分類数として「50」が入力されたことを示している。

【0179】その後、操作者が情報分析作業を完了して、メニュー・バー2603のファイルボタン2605の選択によりポットアップする画面を省略する保存ボタンを押下すると、文書記憶部402により、操作者が指示したファイル名が付与され、文書の情報、分類結果とともに記憶される。

【0180】図27において、分類結果を表示する欄2701にファイルボタン2702を移動させ、ファイルボタンを押下すると、その分類をおこなうに用いた分類に関する情報および分類設定値に関する情報を表示する分類情報表示画面2703が表示される。これにより、用いた設定値の関連づけが容易に把握することができる。

【0181】分類情報表示画面2703には、たとえば、分類に関する情報として分類がおこなわれた日時に関する情報を示す「分類日時」、分類の対象となった文書数に関する情報を示す「分類対象数」等が表示され、また、分類設定値に関する情報として、いくつに分類したかを示す「分類数」、どの品目に基づいて分類をしたかを示す「分類品目」等が表示される。

【0182】分類処理を実行するために新規な表が作成される。図28は、分類結果1を得た後、再度分類処理がおこなわれ、分類結果2が表示された状態を示している。分類結果1を再度表示したい場合は、画面左下部のラベル上の選択領域2801をマウスポインタを移動させ、マウスボタンを押下する。これにより、分類結果1が再度表示される。その後、分類結果2を再度表示させる場合も同様の操作によりおこなうことができる。

【0183】また、図28において、各分類処理の実行に用いた設定値に関する情報が対応する表の所定の表示領域2802に表示される。この表示領域2802は、分類結果の表示を隠さないように表示させることができ、また、その表示位置を移動することもできる。これにより、分類結果と、それに用いた設定値の関連づけが容易に把握できる。

【0184】つぎに、実施の形態1における文書処理装置の文書処理の一連の手順について説明する。図29は、実施の形態1による文書処理装置の文書処理の一連の手順を示すフローチャートである。

【0185】図29のフローチャートにおいて、まず、文書データが文書処理装置に入力されたか否かを判断する（ステップS2901）。ここで、文書データが入力されるのを待つて、文書データが入力された場合（ステップS2901肯定）、入力された文書データを記憶する（ステップS2902）。なお、ステップS2901およびステップS2902の各ステップは、次の入力がある

ことに他のステップとは独自におこなわれるようにしてもよい。

【0186】つぎに、記憶された文データの全部または一部が選択されたか否かを判断する（ステップS2903）。ここで、文書データの全部または一部が選択されるのを待つて、選択された場合（ステップS2903肯定）は、選択された文データの全部または一部の文字列の特徴に関するデータの抽出をおこなう（ステップS2904）。

【0187】その後、ステップS2904において、抽出された文字列の特徴に関するデータに基づいて、分類処理等、所定の加工処理をおこなう（ステップS2905）。続いて、ステップS2905において加工処理がおこなわれたデータを、表形式に展開する等の加工処理をおこなう（ステップS2906）。

【0188】さらに、ステップS2905において加工処理されてデータを元の文書データに関連づけて記憶する（ステップS2907）。また、加工処理の設定値等の加工処理の内容に関するデータも併せて記憶する（ステップS2908）。

【0189】その後、ステップS2905において加工処理されたデータの全部または一部が選択されたか否かを判断し（ステップS2908）、選択されなかった場合（ステップS2908否定）は、ステップS2904へ移行し、以後、ステップS2904～S2909の処理を繰り返すおこなう。一方、ステップS2909において、加工処理されたデータの全部または一部が選択された場合（ステップS2909肯定）は、すべての処理を終了する。

【0190】なお、実施の形態1で説明した文書処理方法は、あらかじめ用意されたプログラムをパーソナルコンピュータやワークステーション等のコンピュータで実行することにより実現される。このプログラムは、ハードディスク、フロッピーディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。またこのプログラムは、上記記録媒体を介して、または伝送媒体として、インターネット等のネットワークを介して配布することができる。

【0191】つぎに、実施の形態2～6に係る情報分類装置について説明する。なお、以下説明する実施の形態2～6においては、上記のように多くのノイズを含んだものであるとの解釈に基づいて、一回の文書集合からの話題（内容）抽出と位置づけ、文分類のためのパラメータ（対象文書集合やクラス数、類似度測定、ストップワード等）を変化させながら複数回の分類を実行させる。その結果を保持・統合する手段を取ること、任意の文書集合にどのような内容が含まれるかを漸次的に収集するものである。

【0192】実施の形態2）この発明の実施の形態2に係る文書分類装置を構成する情報処理システムは、図1に示したように実施の形態1の情報処理システムと同様であるので、その説明は省略する。また、サーバ101およびクライアント102のハードウェア構成についても、図2・図3に示したように実施の形態1と同様であるので、その説明は省略する。

【0193】つぎに、実施の形態2による文書分類装置の機能構成について説明する。図30は、実施の形態2による文 分類装置の構成を概念的に示すブロック図である。

【0194】図300のブロック図において、文書分類装置は、入力部3001と、言語解析部3002と、ベクトル生成部3003と、分類部3004と、分類パラメータ指示部3005と、分類結果記憶部3006と、クラスラスタ特徴表示部3007と、クラスラスタ特徴算出部3008と、分類体系記憶部3009と、クラスラスタ選択指示部3010と、分類体系閲覧操作部3011と、を含む構成である。

【0195】入力部3001、言語解析部3002、ベクトル生成部3003、分類部3004、分類パラメータ指示部3005、分類結果記憶部3006、クラスラスタ特徴表示部3007、クラスラスタ特徴算出部3008、分類体系記憶部3009、クラスラスタ選択指示部3010、分類体系閲覧操作部3011は、ROM202または302、RAM203または303、あるいはディスプレイ装置306またはハードディスク316等の記録媒体に記録されたプログラムに記載された命令にしたがってCPU201または301等が命令処理を実行することにより、各部の機能を実現する。

【0196】ここで、入力部3001は、文書データを入力するものであり、たとえば、キーボード209または311、スキャナ313、OCR機能を備えたスキャナ313、またはネットワーク103を経由して文書や文 群を得ることができる1/F204または309等である。

【0197】また、入力部3001は、上記以外に、文データ取得することのできるものであれば、それらすべてを含む。たとえば、文書データがデータベース化されている場合に、そのデータベースが記録された媒体を本実施の形態の文 分類装置に組み入れた場合も文データの入力とする。

【0198】また、言語解析部3002は、入力部3001により入力された文書データを解析して言語解析情報を得るものであり、ベクトル生成部3003は、言語解析部3002により得られた言語解析情報に基づいて前記文書データに対する文書特徴ベクトルを生成するものである。

【0199】また、分類部3004は、ベクトル生成部3003により生成された文 特徴ベクトル間の類似度

に基づいて文書を分類し、文書の部分集合を生成するものであり、分類パラメータ指示部3005は、分類パラメータを指示するものであり、たとえば、キーボード209または311、マウス210または312、またはネットワーク103を経由して指示情報を得ることができ1/F204または309等である。

【0200】また、分類結果記憶部3006は、分類部3004により生成された結果、すなわち、分類された文書の部分集合に関する情報を記憶するものである。また、クラスラスタ特徴表示部3007は、クラスラスタ特徴算出部3008により算出されたクラスラスタ特徴を表示する。

【0201】クラスラスタ特徴算出部3008は、分類部3004により生成された文書の部分集合の特徴であるクラスラスタ特徴を算出するものである。また、分類体系記憶部3009は、クラスラスタ特徴算出部3008により算出されたクラスラスタ特徴を分類体系の構成要素として記憶するものである。また、分類体系記憶部3009は、クラスラスタ選択指示部3010により選択された文書の部分集合を分類体系の構成要素として記憶するものである。すなわち、クラスラスタ選択指示部3010により選択されたクラスラスタに所属する全ての文書もしくは所属する文書の一部を分類体系の構成要素として記憶するものである。

【0202】クラスラスタ選択指示部3010は、クラスラスタ表示部3007により表示された複数のクラスラスタ特徴の中から所望のクラスラスタを選択するものである。また、クラスラスタ選択指示部3010は、前記分類部3004により生成された文書の部分集合の中から所望の部分集合を選択するものである。また、分類体系閲覧操作部3011は、分類体系記憶部3009に記憶されたデータを閲覧したい場合に、その閲覧の操作をおこなうものである。

【0203】つぎに、文書集合に含まれる話題（内容）を抽出することが重要となる好適な例を、アンテナ調査等により得られた自由記述回答の分析場面を想定し、その具体例を用いて説明する。

【0204】近年、たとえば、インターネット等を利用して短時間に数千〜数万件の自由記述回答を回収することが可能であり、このような機能を用いて大量のテキスト情報の収集をおこなうことができる。

【0205】アンテナ調査により得られた大量のテキスト情報の収集の例として、「オフラインのネットワーク化による無償を享受してください」という質問に対して文書で答えた一つの回答記述を文書とすると、文書集合（クラス）は1件この回答の集合ということになる。

【0206】ここで、操作者（アンテナの分析者）は、そのニーズの一つとして、意見集合（文書集合）にどのような種類の意見（話題）が含まれており、意見の極端を把握したい場合がある。このようなニーズを満たすべく、話題の抽出を類似する意見のまとまり（分類）

により実現し、アンテナ結果にどのような種類の意見が含まれているかを抽出する。

【0207】文書分類は、典型的には大きく分けてつぎの3段階のステップから構成される。第1ステップでは、入力部3001により入力された各文書（意見）について、言語解析部3002が、各文書に含まれる単語（あるいは、特定の連続する文字列）を抽出する。この際、たとえば、形態素解析等の言語解析アルゴリズムを用いる。

【0208】第2ステップでは、抽出された単語を列とし、各文書を行とし、要素を単語の出現頻度とした「単語」×「文書」の行列に基づいて単語で構成される多次空間内に各文書をベクトル表現する。これには、以下の方法があり、本実施の形態においては、すべての方法を実装している。

【0210】（1）行列の列成分をそのまま利用する方法。（2）各文書の長さ（文字の数やページ数等）や分類対象全体の文書集合内での各単語の出現頻度を考慮して重み付けをする方法。（3）上記行列から文書間の内積や主成分分析、数量化理論第3類等を利用しておこなわれる）を適用して潜在的意味空間を構成する方法、等である。

【0211】また、「Representing Documents Using an Explicit Model of Their Similarities」（著者名：Brian T. Bartel, I. Garrison W. Cottrell, and Richard K. Belew、論文名：Journal of the American Society for Information Science、学名名：the American Society for Information Science、ページ：254-271、Vol. 46 No. 4、発行年：1995）において、上記記述的意味空間への取換手法を一般化し、文書間の内積行列に、文書が有するほかの文書への参照情報から生成される共参照情報などを付加した行列を用いて、これらの類似性を反映する空間へ文書や単語を射影するための表現空間変換関数を導出しているものもあり、この方法も利用することができる。

【0212】第3ステップでは、分類部3004が、文特徴ベクトルの類似度を用いて文 を分類する。具体的には分類対象データに対してカイ自乗法的手法、判別分析の方法、クラスタリングの方法等を適用することにより分類が実行される。

【0213】また、類似度としては、内積や余弦、ユークリッド距離、マハラノビスの距離等が考えられ、本実施の形態においては、いずれの方法を用いてもよい。

【0214】また、クラスタリングのアルゴリズムに関してはさまざまなものが公知になっている。クラスタリングは、大別して階層型クラスタリングと非階層型クラスタリングが考えられるが、本実施の形態においては、いずれの方法を用いてもよい。

【0215】また、分類パラメータ指示部3005は、分類部3004が文書特徴ベクトルを分類するための分類パラメータを指示する。分類部3004は、分類パラメータ指示部3005により指示された分類パラメータにしたがって内部に保持される文 特徴ベクトルを分類する。

【0216】このようにして、第1ステップ〜第3ステップの各処理を実行することにより第1回目の文 分類が終了すると、分類結果は分類結果記憶部3006により保持してもよい。

【0217】引き続き、クラスラスタ特徴算出部3008が、分類結果がどのようなクラスラスタを得ることができたのかを示す特徴、すなわちクラスラスタ特徴を算出する。典型的には各クラスラスタに所属する文、あるいはその文の一部を算出するが、その際、クラスラスタの重心との類似度に基づいて文書をソーティングして出力する。

【0218】そのほか、クラスラスタ内で複数の単語、クラスラスタに所属する文書数、クラスラスタ内の文 のばらつきの程度を表すクラスラスタ内の標準偏差のような数値をクラスラスタの特徴を表現するものとして算出する。

【0219】これらのクラスラスタの特徴情報は、操作者に対して出力（表示）されたクラスラスタがどのようなもの（どのような特徴を有するもの）かを把握させるために算出されるものであり、操作者に対してクラスラスタの特徴を示すものであれば、上記の内容（特徴）以外のものであってもよい。

【0220】また、クラスラスタ特徴算出部3008は、上記のようにクラスラスタの特徴を示すもの以外に、クラスラスタ間の関係を示す情報も算出する。階層型クラスタリングの場合は、その上位あるいは下位のクラスラスタを、非階層型クラスタリングの場合は、クラスラスタ重心間の類似度に基づく近接のクラスラスタを算出する。

【0221】つぎに、クラスラスタ特徴表示部3007によるクラスラスタ特徴の表示およびクラスラスタ選択の内容について説明する。図31は、実施の形態2による文 分類装置のクラスラスタ特徴表示部3007の表示の一例を示す説明図である。

【0222】図31において、クラスラスタ単位で操作者ができるようになっており、各クラスラスタは「クラスラ1

D1 欄3101、「メツバ一枚」欄3102、「頻度の高い単語」欄3103、「文書内容」欄3104、「重心との類似度」欄3105等の項目から構成される。

【0223】「クラス1D」欄3101には、クラスタの1Dを示す番号が通し番号で付与され、表示される「メツバ一枚」欄3102はクラスタに所属する文あるいは文の一部の数が出され、表示される。その中で頻度の高い単語が抽出され「頻度の高い単語」欄3103に表示される。「文書内容」欄3104には文の重心が表示され、「重心との類似度」欄3105には、数値化された重心との類似度が表示される。これにより、操作者の理解容易性が向上する。

【0224】操作者は、表示された情報（特徴量）に基づいてクラスタについてその特徴を把握することができ、ここで、内容（特徴）が理解可能なクラスタが一つであれば、操作者はクラスタ選択指示部3010によりクラスタを選択することができる。

【0225】より具体的には、マウス210または312等によって、表示されているクラスタの所定の位置、たとえば、「クラスタ1D」欄3101へカーソル310を移動させ、その位置でクリックすることにより、当該クラスタ1Dのクラスタ全体を選択することができる。なお、選択したクラスタに所属する文書は必ずすべてが選択されるわけではなく、その一部の文書が選択されるようにしてもよい。

【0226】図311においては、「クラスタ1D」欄3101がクリックされ、これにより、クラスタ全体が反転表示しており、当該クラスタ（クラスタ1D「1」）が選択されたことを示している。

【0227】また、操作者は、内容が理解可能であるクラスタが存在しない場合は、分類パラメータ指示部3005により分類パラメータの再設定をおこない、再度分類実行をおこなうことができる。

【0228】クラスタ選択指示部3010により選択されたクラスタ1Dに関するデータは分類体系記憶部3009へ送渡される。分類体系記憶部3009は、このクラスタ1Dに関するデータに基づいてクラスタ特徴算出部3006からクラスタに関する上記特徴量を検索し記憶する。

【0229】また、分類体系記憶部3009は、同様的に、分類結果記憶部3006から分類結果を検索し記憶する。さらに、分類体系記憶部3009は、操作者により入力されたクラスタに関するコメント（たとえば、「ネットワーキングの技術が高い」等）の情報を併せて記憶することもできる。このように、操作者が作成した情報を分類体系の構成要素として記憶することにより、分類体系の利用価値がより向上する。

算や、クラスタ間の意味的な関連を手動であるいは、保持されているクラスタ重心間の類似度等を用いて自動で、構造化・体系化することができる。

【0231】つぎに、実施の形態2の文書分類装置の一連の処理の手順について説明する。図32は、実施の形態2による文書分類装置の一連の処理の手順を示すフローチャートである。図32のフローチャートにおいて、まず、分類の対象となる文書が入力される（ステップS3201）。

【0232】つぎに、入力された文書の書頭が解析され（ステップS3202）、解析された結果、すなわち、抽出された単語に基づいて、文書特徴ベクトルが生成される（ステップS3203）。

【0233】その後、分類パラメータの指示があるのを待って、分類パラメータの指示があった場合（ステップS3204肯定）は、指示があった分類パラメータにしたがって文書を分類し（ステップS3205）、その結果、すなわち、クラスタに関する情報を記憶する（ステップS3206）。

【0234】つぎに、分類されたクラスタの特徴を算出し（ステップS3207）、算出された結果を表示する（ステップS3208）。表示されたクラスタの中から、クラスタが選択されたか否かを判断し（ステップS3209）、選択されなかった場合（ステップS3209否定）は、ステップS3204へ移行し、再度分類パラメータの指示があるのを待つ（ステップS3204）。

【0235】一方、ステップS3209において、クラスタが選択された場合（ステップS3209肯定）は、選択されたクラスタに関して分類体系を生成し、記憶する（ステップS3210）。この際、操作者により入力されたクラスタに関する情報を併せて記憶することもでき、これにより、一連の処理を終了する。

【0236】以上説明したように、実施の形態2による文書分類装置によれば、分類対象である文書群での文書間の類似性に基づいて、各文書をそれら文書間の意味的な関連性を反映しうる表現空間へ変換するための表現空間変換部を算出し、その表現空間で文書分類をおこなうことにより、操作者の意図を反映しうる文書分類を実現することができる。

【0237】したがって、分類部3004によりクラスタを得ることができるとともに、クラスタ特徴算出部3008・分類体系記憶部3009により、クラスタ重心間の類似度等を用いて、クラスタの内部に基づいてクラスタの構造化・体系化をおこなうことができる。

【0238】また、クラスタ選択指示部3010により選択されたクラスタのみを用いて、より操作者の意図したものに近いクラスタの構造化・体系化をおこなうことができる。

【0239】〔実施の形態3〕さて、上述した実施の形

態2に加えて、以下に説明する実施の形態3のように、さらにベクトル記憶部と、ベクトル修正部を含む構成とをするようにしてもよい。

【0240】実施の形態3による文書分類装置を構成する情報処理システムは、図1に示したように実施の形態1と同様であるので、その説明は省略する。また、ヤーバー101およびクワイアット102のハーフトウェア構成についても、図2・図3に示したように実施の形態1と同様であるので、その説明は省略する。

【0241】つぎに、実施の形態3による文書分類装置の機能の構成について説明する。図33は、この部部の実施の形態3による文書分類装置の構成を機能的に示すブロック図である。図33において、実施の形態2の図30と同ーのものに関しては同じ符号を付して、その説明を省略する。

【0242】図33のブロック図において、文書分類装置は、入力部3001、言語解析部3002、ベクトル生成部3003、分類部3004、分類パラメータ指示部3005、分類結果記憶部3006、クラスタ特徴算出部3007、クラスタ特徴算出部3008、分類体系記憶部3009、クラスタ選択指示部3010、分類体系閲覧操作部3011のほか、ベクトル記憶部3301と、ベクトル修正部3302とを含む構成である。

【0243】ベクトル記憶部3301は、ベクトル生成部3003により生成された文書特徴ベクトルを記憶するものである。また、ベクトル修正部3302は、文書特徴ベクトル記憶部3301により記憶された文書特徴ベクトルを、クラスタ選択指示部3010により選択された部分集合に属する文書の文書特徴ベクトルを除去したのこりとなるように修正するものである。

【0244】また、分類部3004は、ベクトル修正部3302により修正された文書特徴ベクトルに基づいて文書を分類する。

【0245】なお、ベクトル記憶部3301、ベクトル修正部3302は、ROM202または302、RAM203または303、あるいはディスク装置306またはハードディスク316等の記憶媒体に記録されたプログラムに記載された命令にしたがってCPU201または301等が命令処理を実行することにより、各部の機能を実現する。

【0246】ベクトル生成部3003において生成された文書特徴ベクトル（列ベクトル）、単語（単語特徴）ベクトル（行ベクトル）はベクトル記憶部3301によって記憶される。これは、次回以降の分類実行の際に利用する文書特徴ベクトルを確保するためである。

【0247】ベクトル修正部3302は、クラスタ選択指示部3010により選択されたクラスタに所属する文書のすべてあるいはその一部の文書を抜き、次回以降もこれらの文書が除かれるよう削除する。削除された文書特徴ベクトルはベクトル記憶部3301により記憶され

る。

【0248】この結果、ベクトル記憶部3301に記憶されているベクトルデータのうち、選択されたクラスタに所属する文書（もしくは操作者に指定されたその一部）列ベクトルを除いたものが、次回以降の分類が実行される際に利用されるデータとなる。

【0249】つぎに、実施の形態3の文 分類装置の一連の処理の手順について説明する。図34は、実施の形態3による文書分類装置の一連の処理の手順を示すフローチャートである。図2のフローチャートにおいて、まず、分類の対象となる文 が入力される（ステップS3401）。

【0250】つぎに、入力された文 の書頭が解析され（ステップS3402）、解析された結果、すなわち、抽出された単語に基づいて、文 特徴ベクトルが生成され（ステップS3403）、生成された文 特徴ベクトルが記憶される（ステップS3404）。

【0251】その後、分類パラメータの指示があるのを待って、分類パラメータの指示があった場合（ステップS3405肯定）は、指示があった分類パラメータにしたがって文書を分類し（ステップS3406）、その結果、すなわち、クラスタに関する情報を記憶する（ステップS3407）。

【0252】つぎに、分類されたクラスタの特徴を算出し（ステップS3408）、算出された結果を表示する（ステップS3409）。表示されたクラスタの中から、クラスタが選択されたか否かを判断し（ステップS3410）、選択されなかった場合（ステップS3410否定）は、ステップS3405へ移行し、再度分類パラメータの指示があるのを待つ（ステップS3405）。

【0253】一方、ステップS3410において、クラスタが選択された場合（ステップS3410肯定）は、選択されたクラスタに関して分類体系を生成し、記憶する（ステップS3411）。この際、操作者により入力されたクラスタに関する情報を併せて記憶することもできる。その後、繰り返し処理をおこなう旨の指示があったか否かを判断する（ステップS3412）。

【0254】ステップS3412において、繰り返しして処理をおこなう旨の指示があった場合（ステップS3412肯定）は、選択されたクラスタに所属する文 のすべてあるいはその一部の文 を除くように文 特徴ベクトルを修正する（ステップS3413）。その後、ステップS3405へ移行し、以後、ステップS3405～S3413の各処理を繰り返しおこなう。

【0255】一方、ステップS3412において、繰り返しして処理をおこなう旨の指示がない場合（ステップS3412否定）は、これにより、一連の処理をすべて終了する。

【0256】以上説明したように、実施の形態3による

文 分類装置によれば、ベクトル修正部33011により、既知になったクラスタの影響を排除した新たなクラスタを生成することができる。

【0257】〔実施の形態4〕さて、上述した実施の形態3においては、ベクトル配座部およびベクトル修正部とを含む構成であったが、以下に説明する実施の形態4のように、ベクトル修正部に代わり、文書表現空間修正部を含む構成とするようにしてもよい。

【0258】実施の形態4による文書分類装置を構成する情報処理システムは、図11に示したように実施の形態1と同様であるので、その説明は省略する。また、サブー101およびクワイアット102のハードウェア構成についても、図2・図3に示したように実施の形態1と同様であるので、その説明は省略する。

【0259】つぎに、実施の形態4による文書分類装置の機能構成について説明する。図35は、この発明の実施の形態4による文 分類装置の構成を機能的に示すブロック図である。図35において、実施の形態2の図30と同一のものに関しては同じ符号を付して、その説明を省略する。

【0260】図35のブロック図において、文書分類装置は、入力部3001、言語解析部3002、ベクトル生成部3003、分類部3004、分類パラメータ指示部3005、分類結果記憶部3006、クラスタ特徴表示部3007、クラスタ選択指示部3008、分類体系記憶部3009、クラスタ選択指示部3010、分類体系閲覧操作部3011のほか、ベクトル記憶部3501と、文 表現空間修正部3502とを含む構成である。【0261】ベクトル記憶部3501は、ベクトル生成部3003により生成された文書特徴ベクトルを記憶するものである。また、文 表現空間修正部3502は、文書特徴ベクトル記憶部3501により記憶された文書特徴ベクトル間の類似度を判断する際、文書表現空間を前記クラスタ選択指示部3010により選択された部分集合から算出する特徴量に基づいて修正するものである。

【0262】また、分類部3004は、文書表現空間修正部3502により修正された文書表現空間を用いて、ベクトル生成部3003により生成された文書特徴ベクトル間の類似度に基づいて文書を分類する。【0263】なお、ベクトル記憶部3501、文書表現空間修正部3502は、ROM202または302、RAM203または303、あるいはディンク装置306またはハードディスク316等の記録媒体に記録されたプログラムに記憶された命令にしたがってCPU201または301等が命令処理を実行することにより、各部の機能を果たす。

【0264】つぎに、文 表現空間修正部3502の内容容について説明する。実施の形態3におけるベクトル修正部3302にあったのは、既知になったクラスタの影響

を排除するために文書特徴ベクトルを除去するが、文書特徴ベクトルを表現する多次元空間自体の変更はおこなわれないう。

【0265】したがって、前回の分類実行の結果、操作者により選択されたクラスタの形成特徴を次の分類実行の際に排除したい場合は、文書ベクトルを表現する空間自体の変更が必要となる。

【0266】そこで、文書表現空間修正部3502を構成し、文書表現空間の修正をおこなうものである。ここで、文書表現空間の特徴次元を変更する例として、操作者により選択されたクラスタの重心と類似度の高い特徴次元の削除をおこなうことについて説明する。

【0267】操作者により選択されたクラスタの重心はベクトルとして表現することができるので、このクラスタ重心ベクトルとベクトル記憶部3501に記憶されている文書表現空間の各特徴次元との類似度を算出することにより、類似度の高い特徴次元を削除する。

【0268】なお、類似度の高い特徴次元を削除する。ユークリッド距離、マハラノビス距離等を用いる。また、判断に関してはある類似度以上を削除対象として採用するようないき値処理による判別や、類似度の高い間にある一定数を削除対象として採用する定数処理による判別を用いる。また、判別分析等も用いることができる。

【0269】文書表現空間修正部3502は、上述のような削除対象の特徴次元を算出して、特徴次元の削除をおこなう。特徴次元の削除は、ベクトル記憶部3501に記憶されている「特徴次元（単語）×「文書」の行列から判別された特徴次元について行ベクトルを削除することによりおこなう。文書表現空間修正部3502により修正された文書ベクトルは、次回以降の分類のために、ベクトル記憶部3501に記憶される。

【0270】つぎに、実施の形態4の文書分類装置の一連の処理の手順について説明する。図36は、実施の形態4による文書分類装置の一連の処理の手順を示すフローチャートである。図36のフローチャートにおいて、まず、分類の対象となる文書が入力される（ステップ3601）。

【0271】つぎに、入力された文書の言語が解析され（ステップ3602）、解析された結果、すなわち、抽出された単語に基づいて、文書特徴ベクトルが生成され（ステップ3603）、生成された文書特徴ベクトルが記憶される（ステップ3604）。

【0272】その後、分類パラメータの指示があるのを待つ。分類パラメータの指示があった場合（ステップ3605肯定）は、指示があった分類パラメータにしたがって文書を分類し（ステップ3606）、その結果、すなわち、クラスタに関する情報を記憶する（ステップ3607）。

【0273】つぎに、分類されたクラスタの特徴を算出

し（ステップ3608）、算出された結果を表示する（ステップ3609）。表示されたクラスタの中から、クラスタが選択されたか否かを判断し（ステップ3610）、選択されなかった場合（ステップ3610否定）は、ステップ3605へ移行し、再度分類パラメータの指示があるのを待つ（ステップ3605）。

【0274】一方、ステップ3610において、クラスタが選択された場合（ステップ3610肯定）は、選択されたクラスタに関して分類体系を生成し、記憶する（ステップ3611）。この際、操作者により入力されたクラスタに関する情報を併せて記憶することもできる。その後、繰り返し処理をおこなう旨の指示があったか否かを判断する（ステップ3612）。

【0275】ステップ3612において、繰り返し処理をおこなう旨の指示があった場合（ステップ3612肯定）は、「特徴次元（単語）×「文書」の行列から判別された特徴次元について行ベクトルを削除することにより文書表現空間を修正する（ステップ3613）。その後、ステップ3605へ移行し、以後、ステップ3605～ステップ3613の各処理を繰り返しおこなう。

【0276】一方、ステップ3612において、繰り返し処理をおこなう旨の指示がなかった場合（ステップ3612否定）は、これにより、一連の処理を終了する。

【0277】以上説明したように、実施の形態4による文書分類装置によれば、前回の分類実行の結果、文書表現空間修正部3502により操作者を選択されたクラスタの形成特徴を次の分類実行時に排除することができ、排除した状態で新たなクラスタを生成することができる。

【0278】〔実施の形態5〕さて、上述した実施の形態3または実施の形態4においては、ベクトル修正部または文書表現空間修正部のいずれか一方のみを含む構成であったが、以下に説明する実施の形態5のように、ベクトル修正部および文書表現空間修正部の両方を含む構成とするようにしてもよい。

【0279】実施の形態5による文書分類装置を構成する情報処理システムは、図11に示したように実施の形態1と同様であるので、その説明は省略する。また、サブー101およびクワイアット102のハードウェア構成についても、図2・図3に示したように実施の形態1と同様であるので、その説明は省略する。

【0280】つぎに、実施の形態5による文書分類装置の機能構成について説明する。図37は、この発明の実施の形態5による文書分類装置の構成を機能的に示すブロック図である。図37において、実施の形態2の図30と同一のものに関しては同じ符号を付して、その説明を省略する。

【0281】図37のブロック図において、文 分類装置は、入力部3001、言語解析部3002、ベクトル生成部3003、分類部3004、分類パラメータ指示部3005、分類結果記憶部3006、クラスタ特徴表示部3007、クラスタ特徴表示部3008、分類体系記憶部3009、クラスタ選択指示部3010、分類体系閲覧操作部3011のほか、ベクトル記憶部3701と、ベクトル修正部3702と、文 表現空間修正部3703とを含む構成である。

【0282】ベクトル記憶部3701は、ベクトル生成部3003により生成された文 特徴ベクトルを記憶するものである。また、ベクトル修正部3702は、文書特徴ベクトル記憶部3701により記憶された文 特徴ベクトルを分類部3004により生成された文 の部分集合の文書特徴ベクトルを除去したのこりの文 特徴ベクトルとなるように修正するものである。

【0283】また、文 表現空間修正部3703は、ベクトル記憶部3701により記憶された文 特徴ベクトル間の類似度を判断する際、文 表現空間を前記クラスタ選択指示部3010により選択されたクラスタ特徴に基づいて修正するものである。

【0284】また、分類部3004は、文 表現空間修正部3703により修正された文書表現空間を用いて、ベクトル修正部3702により修正された文書特徴ベクトル間の類似度に基づいて文 を分類する。

【0285】なお、ベクトル記憶部3701、ベクトル修正部3702、文書表現空間修正部3703は、ROM202または302、RAM203または303、あるいはディンク装置306またはハードディスク316等の記録媒体に記録されたプログラムに記憶された命令にしたがってCPU201または301等が命令処理を実行することにより、各部の機能を果たす。

【0286】つぎに、ベクトル修正部3702および文書表現空間修正部3703の内容容について説明する。実施の形態4においては、選択されたクラスタに所属する文書は次回以降の分類実行の際にも使用される。

【0287】実施の形態5では、ベクトル修正部3702および文書表現空間修正部3703の両方を備えることにより、選択されたクラスタに所属する文 を次の分類実行の際に除去し、次の分類実行の際には分類対象文書としないようにする。

【0288】実施の形態4においては、話題抽出の側面を強調し、ある文書が複数の話題として分類される可能性を前提としており、たとえば、ネットワーク化に関する調査における「エンピューサーがソフトウェアのインストール方法について関心するのデズシステム管理」としての仕事ができない」という回答に対して言えば、これは「ソフトウェアの操作方法に関する困難性」という話題として分類され得るし、「システム管理者の仕事の多忙さ」という話題で分類される可能性もあ

る。
【0288】実施の形態4においては、いずれにしても、「ソフトウエアの操作方法理解に関する困難性」というクラスタと「システム管理者の仕事の多忙さ」というクラスタの両方とも抽出したいというニーズに応えて

いる。
【0290】これとは反対に、操作者は、一度抽出した図解は既知であるので、次の分類の際はなるべく異なる分類結果が欲しいとするケースも考えられる。実施の形態5では、このような要求に応えるため、ベクトル修正部3702により、n回目の分類で選択されたクラスタに所属する文のすべてまたはその一部を次回以降の分類を実行する際、分類対象から除去するものである。
【0291】クラスタ選択指示部3010により選択指示を受けたクラスタの所属文書はベクトル記憶部3701において列ベクトルの形式で記憶されているため、ベクトル修正部3702では劣ベクトルを除去すること

で、次回以降の分類実行用の分類対象文書集合を生成する。
【0292】さらに、実施の形態4と同様に、選択されたクラスタにより文、最現空間修正部3703は、ベクトル記憶部3701に記憶されている行列から特徴次元を削除する。

【0293】つぎに、実施の形態5の文書分類装置の一連の処理の手順について説明する。図38は、実施の形態5による文 分類装置の一連の処理の手順を示すフローチャートである。図38のフローチャートにおいて、まず、分類の対象となる文 が入力される（ステップS3801）。

【0294】つぎに、入力された文書の単語が解析され（ステップS3802）、解析された結果、すなわち、抽出された単語に基づいて、文書特徴ベクトルが生成され（ステップS3803）、生成された文書特徴ベクトルが記憶される（ステップS3804）。

【0295】その後、分類パラメータの指示があるのを待つて、分類パラメータの指示があった場合（ステップS3805肯定）は、指示があった分類パラメータにしたがって文 を分類し（ステップS3806）、その結果、すなわち、クラスタに関する情報を記憶する（ステップS3807）。

【0296】つぎに、分類されたクラスタの特徴を算出し（ステップS3808）、算出された結果を表示する（ステップS3809）。表示されたクラスタの中から、クラスタが選択されたか否かを判断し（ステップS3810）、選択されなかった場合（ステップS3810否定）は、ステップS3805へ移行し、再度分類パラメータの指示があるのを待つ（ステップS3805）。

【0297】一方、ステップS3810において、クラスタが選択された場合（ステップS3810肯定）は、

選択されたクラスタに関して分類体系を生成し、記憶する（ステップ3811）。この際、操作者により入力されたクラスタに関する情報を併せて記憶することもできる。その後、繰り返し処理をおこなう旨の指示があったか否かを判断する（ステップS3812）。

【0298】ステップS3812において、繰り返し処理をおこなう旨の指示があった場合（ステップS3812肯定）は、選択されたクラスタに所属する文書のすべてあるいはその一部の文書を除くように文書特徴ベクトルを修正する（ステップS3813）。

【0299】ステップS3813に引き続き、「特徴次元（単語）×「文書」の行列から判別された特徴次元について行ベクトルを削除することにより文書表現空間を修正する（ステップS3814）。その後、ステップS3805へ移行し、以後、ステップS3805～S3814を繰り返しおこなう。

【0300】一方、ステップS3812において、繰り返し処理をおこなう旨に指示がない場合（ステップS3812否定）は、これにより、一連の処理をすべて終了する。

【0301】以上説明したように、実施の形態5による文書分類装置によれば、ベクトル修正部3702が、既知になつたクラスタの影響を排除し、かつ、文書表現空間修正部3703が、前回の分類実行の結果、操作者に選択されたクラスタの形成特徴を次の分類実行時に排除することができ、排除した状態で新たなクラスタを生成することができる。

【0302】【実施の形態6】さて、上述した実施の形態2または実施の形態4においては、繰り返し分類処理をおこなった場合に、ある文書が何度も選択されたかその情報については考慮していなかったが以下に説明する実施の形態6のように、選択情報付与部を含む構成とし、選択情報をクラスタ特徴とともに表示するようにしてもよい。

【0303】実施の形態6による文書分類装置を構成する情報処理システムは、図1に示したように実施の形態1と同様であるので、その説明は省略する。また、サブパー101およびクライアント102のハードウェア構成については、図2、図3に示したように実施の形態1と同様であるので、その説明は省略する。

【0304】つぎに、実施の形態6による文書分類装置の機能的構成について説明する。図39は、この発明の実施の形態6による文書分類装置の構成を概念的に示すブロック図である。図39において、実施の形態4の図35と同ーのものに同じ記号を付し、その説明を省略する。

【0305】図39のブロック図において、文書分類装置は、入力部3001、言語解析部3002、ベクトル生成部3003、分類部3004、分類パラメータ指示部3006、分類結果記憶部3006、クラスタ特徴表

示部3007、クラスタ特徴算出部3008、分類体系記憶部3009、クラスタ選択指示部3010、分類体系閲覧操作部3011、ベクトル記憶部3501、文書表現空間修正部3502のほか、選択情報付与部3901を含む構成である。

【0306】選択情報付与部3901は、分類部3004により生成された文書の部分集合に所属する文書のすべてあるいは一部が選択された場合に選択されたことを示す選択情報を付与する。また、クラスタ特徴表示部3007は、クラスタ特徴を表示するとともに、選択情報付与部3901により付与された選択情報を表示する。
【0307】なお、選択情報付与部3901は、ROM202または302、RAM203または303、あるいはディスク装置306またはハードディスク316等の記録媒体に記録されたプログラムに記憶された命令にしたがってCPU201または301等が命令処理を実行することにより、機能を実現する。

【0308】つぎに、選択情報付与部3901の詳細な内部について説明する。ブロック図の図40の例において、独自の高いユニークな意見は貴重であることが経験的に知られている。これは、調査を企画する担当者が予想できなかった意見である場合が多いからである。

【0309】そこで、操作者に選択されたクラスタに所属する文書を、次回以降の分類実行の際に使用する場合において、クラスタ特徴表示部3007で過去の文書を表示する際に、各文書が何回選択されたかを示すことで、多量に利用される文書の識別性を向上させ、かつ一度も選択されない文書の識別性も向上させることができる。

【0310】図40は、実施の形態6による文書分類装置の分類結果記憶部3006において収められたテーブル4000を示す説明図である。図40において、文書IDごとにテーブル化されており、テーブル4000は、各文書がどのサイクルに分類実行の際に操作者に選択されたかを記録する。すなわち、選択された場合は選択情報として「1」を記録し、選択されなかった場合は選択情報として「0」を記録する。

【0311】たとえば、4回分類が実行された際、文書IDの「1」、第1回目および第2回目の分類実行時に操作者に選択されたことを示し、第3回目、第4回目の分類実行時には選択されなかったことを示している。一方、文書IDの「2」は、未だ一度も選択されておらず、操作者によって未知の意見という可能性を示唆している。

【0312】こうした情報に基づいて、クラスタ特徴表示部3007が文書を操作者に表示する際、たとえば、選択された回数に応じて表示を変化させるようにするとよい。変化する速度や程度等が考えられる。

【0313】また、直接的に数字やグラフ等で選択され

た回数を強調することもできる。いずれにしても選択される文書と一度も選択されていない文 とを視覚的に識別できる表示形式であれば、上記のものに限らない。
【0314】また、上記選択情報を分類体系閲覧操作部3011の閲覧操作により閲覧できるようにしてもよい。

【0315】つぎに、選択情報付与部3901の処理の内容について説明する。図41は、実施の形態6による文書分類装置の選択情報付与部3901の処理の手順を示すフローチャートである。図41のフローチャートにおいて、まず、分類処理がおこなわれ（ステップS4101）、それに引き続き、最初の文 が抽出される（ステップS4102）。

【0316】抽出された文 が、ステップS4101における分類処理の際に選択されたか否かを判断する（ステップS4103）。ここで、選択された場合（ステップS4103肯定）は、選択情報としてデータ「1」を記録する（ステップS4104）。一方、選択されなかった場合（ステップS4103否定）は、選択情報としてデータ「0」を記録する（ステップS4105）。

【0317】つぎに、すべての文 について処理が終了したか否かを判断する（ステップS4106）。ここで、すべての文書について処理が終了していない場合（ステップS4106否定）は、つぎに文書を抽出し（ステップS4107）、ステップS4103へ移行し、以後、ステップS4103～S4107を繰り返しおこなう。

【0318】一方、ステップS4106において、すべての文書について処理が終了した場合（ステップS4106肯定）は、ステップS4101へ移行し、再度分類処理がおこなわれる（ステップS4101）。このようにして、分類処理がおこなわれる回数だけ、ステップS4101～S4107の各処理が繰り返しおこなわれる。

【0319】以上説明したように、実施の形態6によれば、選択情報付与部3901が選択情報を付与し、その選択情報をクラスタ特徴表示部3007が表示するので、多量に利用される文 の識別性および一度も選択されない文書の識別性を向上させることができる。

【0320】なお、実施の形態2～5で説明した文 分類方法は、あらかじめ用意されたプログラムをパーソナルコンピュータやワークステーション等のコンピュータで実行することにより実現される。このプログラムは、ハードディスク、フロッピーディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。またこのプログラムは、上記記録媒体を介して、または伝送媒体として、インターネット等のネットワークを介して配布することができる。

【0321】つぎに、実施の形態7～16に係る情報分類装置について説明する。本発明の実施の形態では、自然言語で記述された一つ以上の文の集まりであり、かつその一つ以上の文の集まりが分類される対象である場合、それを文と扱う。具体的な例をあげれば、1PC分類等により分類される公開特許公報や、政治・経済・文化・科学技術分野の特定分野に分類される新聞記事も文であるし、それから請求項や特定の一文を取り出したものであっても、請求項という分類に含まれる文であるが、用途等により分類可能な特定の一文であれば文書とみなす。以下、図面によりこの発明の実施の形態7～16を詳細に説明する。

【0322】〔実施の形態7〕図42はこの発明の実施の形態7を示す文 分類装置の構成ブロック図である。図42に示したように、実施の形態7の文書分類装置は、文 データ群を入力する文書入力部（文書入力手段）5001、それぞれの文書データを所定の基準に基づいて一つまたは複数の分類文書データに分類する文書分割部（文 分割手段）5002、上記文書データと分類文 データとを対応付けるマップを生成する文書一分割文 対応マップ生成部（文書一分割文書対応マップ生成手段）5003を備えている。

【0323】また、上記文書分類装置は、分類文書データつまり分割された文 を分類する分類文書分類部（分類文 分類手段）5004、分類文書分類結果情報を生成する分類文 分類結果生成部（分類文書分類結果生成手段）5005、上記文 一分割文書対応マップと上記分類文 分類結果情報とを用いて上記文書データの分類結果情報とを生成する 分類結果生成部（文書分類結果生成手段）5006などを備えている。

【0324】なお、上記文書分割部5002、文書一分割文 対応マップ生成部5003、分類文書分類部5004、分類文 分類結果生成部5005、文書分類結果生成部5006は共有または独自のプログラム記憶用メモリおよびプログラムにしたがって動作するCPUを有している。

【0325】以下、図42などにしたがって、実施の形態7の文 分類装置、文 分類方法を詳細に説明する。まず、文 入力部5001により、文書群が入力される。上記文 入力部5001はキーボード、OCR装置、着脱型記録媒体、またはネットワーク通信手段を備え、それらのいずれか一つを介して文書データ群を入力するのである。

【0326】そして、文 分割部5002が上記文書データ群を取得し、それぞれの文書データを所定の基準に基づいて分割し、一つの文書データから一つまたは複数の分類文 データを生成する。なお、文書データを分割する方法としては、文 の構造情報や文書を構成する要素情報を用いたり、利用者が指定する方法などを用いるが、ここでは、その方法は問わないこととする。

【0327】図43に、この文書分類装置／文書分類方法でおこなわれる、文書データから複数の分類文書データを生成する一例を示す。この例に示した文書1には複数のニュースストロップが記述されており、1日分のストロップが文書単位となっている。図示したように、この文書ではそれぞれのニュースストロップが二つの改行コードにより分離されているので、この規則を用いて一つの文書である文書1を分割し、一つが一つのストロップにより形成される分類文書1～1～7の7つの分類文書データを生成する。なお、分割前の文書1も分類文書データとして含めることもできるが、ここでは含めないことにする。

【0328】文書が分割されると、文書一分割文書対応マップ生成部5003が分割前の文書データとその文書データから生成された分類文書データとを対応付けるマップを生成する。たとえば、個々の文書データを一意に示す識別子と個々の分類文書データを一意に示す識別子とから構成されるマップ、あるいは文書データごとに分割文書データを一意に示す識別子からなるマップを生成するのである。なお、文書データと分類文書データを対応付ける方法についてはここでは問わないこととする。

【0329】図44に、文書一分割文書対応マップを生成する一例を示す。図44において、文書1～文書3は文書データを示し、分類文書～分割文書12は分類文書データを示している。図示のように、それぞれの文書データおよび分割文書データにそれぞれを一意に識別することのできる識別番号（識別子）を付与し、上記文書データの識別番号と分割文書データの識別番号とを図44の左下に示したテーブル形式で対応づけている。なお、任意の複数の分類文書データが文書分類にて用いられる基準において同一とみなすことができる場合は、それらの識別番号を同一にしてもよい。

【0330】続いて、分類文書分類部5004が上記分割文書を対象に文書分類をおこなう。個々の分類文書に対して、たとえば、言語処理を施し、文書中に含まれているそれぞれの単語の出現頻度を計数し、それに基づいてそれぞれの文書の特徴を計量的に表す特徴ベクトルを求め、それらの特徴ベクトルに対してカイ自乗法、判別解析手法、またはクラスタ解析法などを適用することにより文書分類をおこなう。

【0331】つぎに、図45に示すように、分類文書分類結果生成部5005が上記の分類文書分類の結果に基づいた分類文書分類結果情報を生成する。

【0332】ここで、分類文書分類結果情報とは、たとえば、各分割文書データの所属カテゴリに関する情報（たとえば、図45に示した「分類文書データを3つのカテゴリに分類した結果」という表中の「分類カテゴリ」および「所属カテゴリの代表値との距離」の項の情報）と、生成された所属カテゴリ図々に関する情報（たとえば、図45に示した「分類カテゴリに関する情報」と

いう表中の「代表値」および「所属データ数（分割文書数）」の項の情報）、生成された所属カテゴリ間の情報（たとえば図45に示した「分類カテゴリ間の距離」という表の中間の数）などである。なお、利用者は上記のような種々の情報を分類結果分析の基礎データとして利用することができる。

【0333】図45は、12個の分割文書データをそれらの有する計量的特徴ベクトルを用いて3つのカテゴリに分類した場合の分類結果の生成例である。分割文書データの有する計量的な3次元ベクトル（ベクトルの成分数は分類対象文書群に生成するすべての単語の種類数になるが、ここでは、いくつかの単語が精選した3次元ベクトルに線形変換している）に対してたとえばクラスタ分析手法の一つであるWarf法などを適用することによって3つのカテゴリに分類することができる。

【0334】つまり、各分割文書データは図示したように3つのカテゴリのうちのいずれか一つに属する。なお、所属カテゴリの代表値とは、所属分割文書データの特徵ベクトルの平均値（所属分割文書データの重心）である。

【0335】また、所属カテゴリの代表値との距離（類似度に対応する）は、たとえば、図45の分割文書3に示している、分割文書データ特徴ベクトルの項における分割文書3の値と、分割文書3の分類カテゴリであるカテゴリ2の代表値（所属分割文書データの重心）の項の値により、以下の数式から求めることができる。

【0336】 $\left((3.00 - 2.66)^2 + (2.00 - 2.00)^2 \right)^{1/2} = (4.00 - 3.66)^2)^{1/2} = 0.48$

上記の所属カテゴリの代表値との距離が小さいほど、そのカテゴリに属する平均的分割文書との類似度が高いということになる。

【0337】なお、分割文書分類結果情報としては、図45に示した以外にも、カテゴリ内分散やカテゴリ間分散、各カテゴリにおける類似度のレンジなどさまざまな統計量を生成することができる。

【0338】続いて、文書分類結果生成部5006が上記文書一分割文書対応マップと上記分割文書分類結果情報とを用いて、たとえば図46に示すような、上記文書データの分類結果情報を生成する。図46の例では、図示したように、各分類カテゴリごとに、所属する分割文書データ、その類似度（所属カテゴリの代表値との距離）、分割文書データの属する分割前文書データ（所属文書）、文書占有率（分割文書データの当該カテゴリに所属する割合）、分割文書データの所属文書における相対位置（順序）、所属カテゴリ内での当該分割文書データの類似度の順位などを生成している。

【0339】なお、上記において、所属文書は文書一分割文書対応マップから、それ以外の分類結果情報は分割文 分類結果情報から得ている。文 分類結果生成部5006は図46に示した情報以外にも、各カテゴリ内での分散、分割文書データの所属カテゴリ内での位置などさまざまな統計量、文書データや分割文 データの内容とを基に分類結果情報として利用することもできる。

【0340】また、上記においては、すべての結果を分割文書データを単位とした表形式で表現しているが、分類カテゴリや文書データを単位として表現することもできる。また、分類結果情報をテキスト表現するだけでなく、グラフィカルな表現にて、利用者が理解しやすいようにすることも可能である。

【0341】こうして、本実施の形態によれば、一つの文書が分割され、分割文書が分類され、分割前文 と上記分割文書との対応が利用者に示され、上記分割文書の分類結果が利用者に示されるので、一つの文 の中に複数の話題や意味が含まれている場合に、ある特定の話題や意味に限定されたカテゴリに分類されたり、利用者の意図するカテゴリとは異なるカテゴリに分類されたりすることがなく、したがって、利用者がその分類カテゴリをよく理解できる。また、分割前文（所属文書）中の分割文書の位置なども示されるので、利用者は文 群中の読みたい部分を効率的に読むことができる。

【0342】〔実施の形態8〕図47は本発明の実施の形態8に係る文書分類装置の構成ブロック図である。図示したように、実施の形態8の文 分類装置は、図42に示した実施の形態7の構成に加え、文 データを保存する文書保存部（文書保存手段）5007、分割文書データを保存する分割文 保存部（分割文書保存手段）5008、文書一分割文 対応マップ生成部5003により生成された文書一分割文書対応マップを保存する文書一分割文書対応マップ保存部（文 一分割文 対応マップ保存手段）5009を備えている。なお、上記各保存部にはたとえば共有のハードディスクや半導体メモリなどにより構成される。

【0343】上記した構成により、本実施の形態の文 保存部5007は、文 データの内容や、文 の作成者、作成日、最終修正日などの文 データに付随する情報を適切な形式で保存する。また、文 データが文 内容とともにその要素からなる計量的な特徴ベクトルを持つ場合にはこれらも保存する。文 入力部5001にて、個々の文書データにそれらを一意に表す識別子が付与される場合にはこの識別子も適切な形式で保存することができる。

【0344】また、分割文 保存部5008は、文 分割部5002により生成される分割文 データの内容を適切な形式で保存するとともに、計量的な特徴ベクトルを持つ場合にはこれらも保存する。個々の上記分割文書データにそれらを一意に表す識別子が付与される場合にはこの識別子も適切な形式で保存することができる。

【0345】また、文 一分割文 対応マップ保存部5009は、文 一分割文 対応マップ生成部5003に

より生成される文 一分割文書対応マツタを適切な形式で保存する。

【0346】このように、実施の形態8によれば、文書データ、分割文 データ、および文書一分割文書対応マツタが保存されるので、分割文書データおよび文書一分割文 対応マツタを再生生成することになり、同一の文書データに対して、分類数、分類手法、または分類時の設定などパラメータの異なる分類結果を効率的に求めることができる。また、文 データを分類し、分類結果を生成するために必要なデータが保存されることにより、利用者は、分類作業に対して時間的な自由度を持つことができ、過去に行った文 分割の再分析を任意の時間におこなうこともできる。

【0347】〔実施の形態9〕図48は本発明の実施の形態9を示す文 分類装置の構成ブロック図である。図48に示したように、本実施の形態の文書分類装置は、図47に示した実施の形態8の構成に加え、分割文書分類結果生成部5005により生成された分割文書分類結果情報を保存する分割文 分類結果保存部(分割文書分類結果保存手段)5010を備えている。なお、上記分割文 分類結果保存部5010は、たとえば、共有のハードディスクや半導体メモリなどにより構成される。

【0348】このように、第3の実施の形態によれば、文 データ、分割文 データ、文書一分割文書対応マツタ、および、分割文書分類結果情報が保存されるので、実施の形態8の効果に加え、一度分割を実行すれば、その分類結果をテキスト表現や表表現やグラフ表現などさまざまな形式で表現することができ、また、分割文書分類結果情報が保存されることにより、分割の実行作業および分割結果の分析作業において、利用者は、時間的な自由度を持つことができ、過去に行った文書分類結果の再分析をさまざまな表現形式で任意の時間におこなうこともできる。

【0349】〔実施の形態10〕この発明の実施の形態10では、前記各実施の形態の文書分類装置、文書分類方法において、図49に示すように、文書分割部5002により生成される複数の分割文書データ中に分割前の文 データである文書1を含む。これにより、本実施の形態では、利用者は、分割されている文書データを分類することによって得られる詳細な文書データの分類構造だけでなく、分割前の文 データ自体を分類した結果として得られるマクロな分類構造の融合した分類構造を得ることができ、

【0350】〔実施の形態11〕この発明の実施の形態11では、前記各実施の形態の文書分類装置、文書分類方法において、文 分割部5002は、文書データの構造情報を基に文 データを分割する。図50に、分類対象文 データがHTML形式で記述された文書の例を示す。分割をおこなう前に、図50に示したようなHTML形式の文 データから構造情報を抽出し、それらの構

造を用いて文書の適切な分割規則を設定することにより文書データから分割文書データを生成する。

【0351】つまり、この例では、文書データ中のタグに着目し、「タグ」を持つテキストを一つの分割文書データとする」という文書を分割マツタを生成する規則とする。この規則を文書データに適用することにより図50に示したような7つの分割文書が生成される。

【0352】上記のように、文書が、HTML、XML、SGMLなど特定の構造に文書の形式を有していない場合でも、文字の大きさ、文字の装飾、文字の色、およびフォントなどに関する情報から分割規則を生成し、分割文書を生成することもできる。また、文書データがイメージであってOCR装置などにより入力される場合には、元のイメージのレイアウト情報などを利用することにより分割規則を生成し、分割文書を生成することもできる。

【0353】なお、文書データのすべてをいずれかの分割文書データにする必要はない。たとえば、図50に示した例では、文字列「ニューストピック(98/09/25)」は分割文書には採用しない。

【0354】このように、実施の形態11では、文書データから構造情報を抽出し、文書分割をおこなう前に構造情報を用いて文書の適切な分割規則を設定することにより、異なった話題の適切な分割をおこなうことができ、したがって、文書データの詳細な分類構造がわかる文書分類を適切におこなうことができる。

【0355】〔実施の形態12〕この発明の実施の形態12では、前記実施の形態7〜10の文書分類装置、文書分類方法において、図51に示すように、文書データに含まれる単語など要素を抽出する文書要素解析部(文書要素抽出手段)5011、上記文書要素解析部5011により抽出された要素に付随する品詞など要素付随情報を抽出する要素付随情報抽出部(要素付随情報抽出手段)5012を備え(図51は図48に示した実施の形態9に文書要素抽出部5011、要素付随情報抽出部5012を加えた例で示している)、文書分割部5002が、上記文書要素解析部5011により抽出された要素、または上記要素と上記要素付随情報抽出部5012により抽出された要素付随情報とを用いて上記文書データを分割する。

【0356】図52に示すように、文書分割をおこなう前に、自然言語処理手段である文書要素解析部5011が文書データから単語などそれらの要素を抽出し、要素付随情報抽出部5012が品詞など要素付随情報を抽出して文書の適切な分割規則を設定するのである。なお、上記文書要素解析部5011および要素付随情報抽出部5012は新たに設けられるのではなく、分割文書分類部5004内の同様の手段を用いることが可能である。

【0357】この実施の形態では、たとえば、図52に

示したように、文書データが特定の構造情報を持たない複数のニューストピックの集まりであり、各トピックが、単語「トピック」+「数字」+「改行コード」という文字列の後に記述されている場合で説明すると、上記のような構造が文書要素解析部5011および要素付随情報抽出部5012の抽出結果から認識され、文章の終端を考慮して、「トピック+数字+改行コード」という文字列を先頭とし、上記文字列または文書終端記号を終端として囲まれる文字列を一つの分割文書データとしてという分割規則が生成されることになる。

【0358】さらに詳しく説明すると、抽出された単語とその品詞情報などから、まず、名詞と改行コードのみを抽出し、つぎに、文字列「トピック+数字+改行コード」および文書終端記号を抽出し、文書内でそれらの位置を記憶する。そして、文書データに対して前記分割規則を適用し、図52に示したような分割文書データを生成する。

【0359】なお、文書データのすべてをいずれかの分割文書データにする必要はなく、たとえば、図52に示した例では、文字列「ニューストピック(98/09/25)」は分割文書には採用しない。また、上記の例では、文書データから要素およびその付随情報を抽出して分割規則を設定する場合で説明したが、要素のみを抽出してその要素情報から分割規則を設定することも可能である。

【0360】こうして、実施の形態12によれば、文書データからそれらの要素情報などを抽出し、抽出した要素情報などを用いて文書の分割規則を設定することにより、実施の形態11と同様に、文書データの詳細な分類構造がわかる文書分類を適切におこなうことができる。

【0361】〔実施の形態13〕この発明の実施の形態13では、前記実施の形態7〜10の文書分類装置、文書分類方法において、利用者がより指示された指定範囲にしたがって文書分割部5002が文書データを分割する。図53に示すような文書データに対して利用者がそれぞれ分割文書の範囲を指定すると、指定にしたがって文書分割部5002が文書分割をおこなう。

【0362】本実施の形態では、文書分割時、文書分割部5002がまず、画面上に、その初期状態として左右の指示ポイントおよび領域指定ラインからなる領域指定オブジェクトを文書の最上部に提示する。この状態で、利用者は、マウスなどポインティングデバイスを用いて、左右どちらかの指示ポイントをクリックし、それを上下に移動させることにより、それぞれの分割文書の領域を選択することができる。

【0363】また、この指定時、文書分割部5002は、領域選択処理をおこなっていることを示すため、指示ポイントと領域を黒色から白色に、領域指定ラインを黄緑から破線に変化させる。選択領域を決定するには、希望の位置で指示ポイントのドラッグを止めればよい。

【0364】つぎに、利用者は選択した領域を分割文とは、それを明示的に表示するために、文 分割部5002は選択領域を図示のように黒付け表示にさせる。

【0365】こうして、本実施の形態によれば、利用者は文書データからそれぞれの分割文 データを所望通りに選択することができるので、文 データの詳細な分類構造がわかり、かつ利用者の意図に合った文 分割をおこなうことができる。

【0366】〔実施の形態14〕この発明の実施の形態14では、前記実施の形態7〜10の文 分類装置、文書分類方法において、文 データ中の文字数、文数、または文字数と文数の両方を基に文 データを分割する。たとえば、図54に示す文 データをほぼ500文字を単位として分割をおこなう。

【0367】ここで、ほぼ200文字を単位とするのは、正確な200文字単位としてもその終端が句点である保証がないことから、200文字目の前後のもつとも近い句点をそれぞれの分割文 の終端とするからである。こうして、図54に示したような分割文書が生成される。同様に、所定の文数を単位とした文 分割をおこなうこともできるし、文字数と文数の両方を基にした文書分割をおこなうこともできる。

【0368】このように、実施の形態14によれば、文字数、文数、または文字数と文数の両方を基に文 データを分割することにより、話題の異なった内容などが異なった分割文書として分割され、分類される可能性が高くなるので、文書データの詳細な分類構造がわかる文 分割をおこなうことができる。

【0369】〔実施の形態15〕この発明の実施の形態15では、前記各実施の形態の文 分類装置、文 分類方法において、文書分類結果生成部5006が分類結果情報として、文書データを示す情報および上記文 データに付随する代表的情報のみを提示する。

【0370】たとえば図55に示すように、先頭に分類キーワードを表示し、その後にそのキーワードを代表するキーワードを表示し、キーワードの下には文書データを示す情報として当該キーワードに属する分割文 データを含んでいる文書データの、たとえば、文 データ名(文書名)を表示する。また、各 データ名の左側には文書アイコンを表示させ、この文 アイコンが指示されたとき、文書データの内容を表示させる。

【0371】また、各文 データ名の配置順は、キーワード代表値との類似度が高い分割文 データの文 データ名を先(左側)にする。また、同じ文 データから生成された複数の分割文 データが同一の分類キーワードに属している場合には、類似度のもっと高い分割文 データに対応する文書データの名を表示する。なお、上記キーワードとは出現頻度の多い単語である。

【0372】このように、実施の形態15によれば、文

分類結果が文 データを示す情報と文書データに付随する代表的情報のみが表示されるので、利用者は文書データの詳細な分類構造の概要を容易に把握することができ

る。

【0373】(実施の形態16) この発明の実施の形態16では、実施の形態15の文書分類結果提示に加えて、分類文 データを示す情報および上記分類文データに付随する情報を提示する。

【0374】たとえば、図56に示すように、先頭に分類カテゴリ名を表示し、その横にそのカテゴリを代表するキーワードを表示し、カテゴリ名の下には文書データを示す情報として当該カテゴリに属する分類文書データをきんている文 データのたとえば文書データ名(文書名)を表示する。

【0375】また、各文 データ名の左側には文書アイコンを表示せ、この文 アイコンが指示されたとき、文 データの内容を表示させる。また、文書データ名の右側には分類文 アイコンを表示させる。なお、分類文 アイコン中には当該文書データにおける分類文書データの位置と当該文 データ中の分類文書数を表示させる。さらに、上記分類文 アイコンを指示することで文 データ中の当該分類文 データを表示させることができる。

【0376】また、各文 データ名の配置場所はカテゴリ代表値との類似度が高い分類文書データの文書データ名を先にする。また、同じ文書データから生成された複数の分類文 データが同一の分類カテゴリに属している場合には類似度の順位がわかるようにその順位を表示させる。

【0377】このように、実施の形態16によれば、文 分類結果が文書データを示す情報と文書データに付随する代表的情報、および分類文書データを示す情報と分類文 データに付随する代表的情報のみが表示されるので、利用者は文 データの詳細な分類構造の概要ととも

にどの分類文 が起因して当該カテゴリに分類されたかというようにすることも容易にわかる。

【0378】以上、本発明の文書分類装置および文書分類方法を説明したが、この文書分類方法を実現するプログラムを着脱可能であるとともにコンピュータ読み取り可能な記録媒体に記録し、上記記録媒体を移した先の情報処理装置内で本発明によった文書分類をおこなうこともできる。

【0379】

【発明の効果】以上説明したように、請求項1の発明によれば、入力された文 データを記憶する文書記憶手段と、前記文 記憶手段により記憶された文書データの全部または一部を選択する選択手段と、前記選択手段により選択された文 データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出手段と、前記特徴抽出手段により抽出された文字列の特徴に関するデー

タに基づいて前記文書データの全部または一部を加工処理する加工処理手段と、前記加工処理手段により加工処理された文書データの全部または一部を出力する出力手段とを備えるため、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

【0380】また、請求項2の発明によれば、前記出力手段が、前記加工処理手段により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定手段と、前記項目値設定手段により設定された項目値ごとに前記文書データの全部または一部を属する属化手段と、を備え、前記文書データの全部または一部を、項目値を少なくとも一つの軸とする数形式に展開して出力するため、簡易な操作で加工処理の結果をクロス渡として表すことができ、情報の内容の把握を容易におこなうことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

【0381】また、請求項3の発明によれば、前記出力手段が、さらに、前記加工処理手段により加工処理された文書データの全部または一部を、前記加工処理手段により加工処理される前の文書データの全部または一部とともに出力するため、加工処理すべき対象データとその他のデータが同時に表示され、それを確認することにより、加工処理の対象範囲の決定を正確かつ容易におこなうことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

【0382】また、請求項4の発明によれば、前記文書記憶手段が、さらに、前記加工処理手段により加工処理された文書データの全部または一部を記憶するため、以後、他の文書データと同様に扱うことができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

【0383】また、請求項5の発明によれば、前記選択手段が、さらに、前記出力手段により出力された文書データの全部または一部を選択するため、出力手段により出力された文書データの全部または一部をさらなる分析の対象とすることができ、多段階で高度な情報分析作業ができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

【0384】また、請求項6の発明によれば、前記文

記憶手段が、さらに、前記加工処理の内容に関するデータの紛失を防止し、当該データの管理が容易になるだけでなく、加工処理に用いた設定とそれによる処理結果を関連づけて把握することができることから、文書の意味に係わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理装置が得られるという効果を奏する。

【0385】また、請求項7の発明によれば、入力手段が、文書データを入力し、言語解析手段が、前記入力手段により入力された文書データを解析して言語解析情報を得、ベクトル生成手段が、前記言語解析手段により得られた言語解析情報に基づいて前記文書データに対する文書特徴ベクトルを生成し、分類手段が、前記ベクトル生成手段により生成された文書特徴ベクトルの類似度に基づいて文書进行分类し、文書の部分集合を生成し、クラス特徴算出手段が、前記分類手段により生成された文書の部分集合の特徴であるクラス特徴を算出し、分類体系記憶手段が、前記クラス特徴算出手段により算出されたクラス特徴を分類体系の構成要素として記憶するため、クラス特徴を得ることができるとともに、クラス重心間の類似度等を用いて、クラスタの内容に基づくクラスタ間の構造化・体系化をおこなうことができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類装置が得られるという効果を奏する。

【0386】また、請求項8の発明によれば、入力手段が、文書データを入力し、言語解析手段が、前記入力手段により入力された文書データを解析して言語解析情報を得、ベクトル生成手段が、前記言語解析手段により得られた言語解析情報に基づいて前記文書データに対する文書特徴ベクトルを生成し、分類手段が、前記ベクトル生成手段により生成された文書特徴ベクトル間の類似度に基づいて文書进行分类し、文書の部分集合を生成し、クラス特徴算出手段が、前記分類手段により生成された文書の部分集合の特徴であるクラス特徴を算出し、表示手段が、前記クラス特徴算出手段により算出された文書の部分集合の特徴であるクラス特徴を算出したクラスタ特徴を表示し、クラスタ選択指示手段が、前記分類手段により生成された文書の部分集合の中から所望の部分集合を選択し、分類体系記憶手段が、前記クラスタ選択指示手段により選択された文書の部分集合を分類体系の構成要素として記憶するため、選択されたクラスタの構造化・体系化をおこなうことができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類装置が得られるという効果を奏する。

【0387】また、請求項9の発明によれば、請求項8の発明において、文 特徴ベクトル記憶手段が、前記ベ

クトル生成手段により生成された文 特徴ベクトルを記憶し、ベクトル修正手段が、前記文 特徴ベクトル記憶手段により記憶された文書特徴ベクトルを、前記クラスタ選択指示手段により選択された部分集合に属する文の文書特徴ベクトルを除去したのりとなるように修正し、前記分類手段が、前記ベクトル修正手段により修正された文書特徴ベクトルに基づいて文 を分類するため、既知になったクラスタの影響を排除した新たなクラスタを生成することができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類装置が得られるという効果を奏する。

【0388】また、請求項10の発明によれば、請求項8の発明において、文 特徴ベクトル記憶手段が、前記ベクトル生成手段により生成された文 特徴ベクトルを記憶し、文書表現空間修正手段が、前記文 特徴ベクトル記憶手段により記憶された文 特徴ベクトル間の類似度を判断する際、文書表現空間を前記クラスタ選択指示手段により選択された部分集合から算出する特徴量に基づいて修正し、前記分類手段が、前記文 表現空間修正手段により修正された文 表現空間を用いて、前記ベクトル生成手段により生成された文 特徴ベクトル間の類似度に基づいて文書进行分类するため、前回の分類実行の結果、操作者に選択されたクラスタの形成特徴を次の分類実行時に排除することができ、排除した状態で新たなクラスタを生成することができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類装置が得られるという効果を奏する。

【0389】また、請求項11の発明によれば、請求項9の発明において、文書特徴ベクトル記憶手段が、前記ベクトル生成手段により生成された文書特徴ベクトルを記憶し、文書表現空間修正手段が、前記文 特徴ベクトル記憶手段により記憶された文 特徴ベクトル間の類似度を判断する際、文書表現空間を前記クラスタ選択指示手段により選択されたクラスタ特徴に基づいて修正し、前記分類手段が、前記文書表現空間修正手段により修正された文書特徴ベクトルを用いて、前記ベクトル修正手段により修正された文書特徴ベクトル間の類似度に基づいて文書进行分类するため、既知になったクラスタの影響を排除し、かつ、前回の分類実行の結果、操作者に選択されたクラスタの形成特徴を次の分類実行時に排除することができ、排除した状態で新たなクラスタを生成すること

ことができ、これにより、任意の文 集合にどのような内容が含まれるかを漸次的に収集することが可能な文 分類装置が得られるという効果を奏する。

【0390】また、請求項12の発明によれば、請求項8または10の発明において、選択情報与手段が、前記分類手段により生成された文 の部分集合に所属する文 のすべてあるいは一部が選択された場合に選択され

たことを示す選択情報を与え、前に表示手段が、前記ラウエー特徴を表示するとともに、選択情報付手段により付与した選択情報と表示するため、多重に利用される文の識別および一度も選択されない文書の識別性を向上させることができ、これにより、任意の文書集合にどのような内容が含まれるかを効率的に収集することと可能な文書分類装置が得られるという効果を奏す。

【0399】また、請求項16の発明によれば、請求項

【0391】また、請求項13の発明によれば、請求項8～12の発明において、前記分類体系を記憶する手段が、前記選択指示手段により選択された文の部分集合に属する全部あるいは一部の文のほか、クラス分特微および／または操作者が作成した任意の情報も分類体系の構成要素として記憶するため、クラスその内容把握を容易にし、かつ、操作者独自の分類体系を簡単に生成できるのにより、任意の文集合にどのような内容が含まれるかを体系的に収集することが可能な文書分類装置が得られるという効果を奏する。

15の発明において、前記分類文書分類結果生成手段により生成された分類文書分類結果情報を保存する分類文書分類結果保存手段を備えるため、請求項15の発明の効果に加え、一度分類を実行すれば、その分類結果をテキスト表現や数値表現やグラフ表現などさまざまな形式で表現することが可能で、また、分割文書分類結果情報を保存することにより、分類の実行作業および分類結果の分析作業において、利用者が時間的な自由度を持つことが可能で、過去に行った文書分類結果の再分析をさまざまな表現形式で任意の時間におこなうことも可能な文書分類装置が得られるという効果を奏する。

【03995】また、請求項17の発明によれば、請求項14～16の発明において、前記文書分類手段により生成される複数の前記文書データには、分割されている文書データを分類することで得られる詳細な文書データの分類と、そのものをさらに、利用者は、分割されている文書データを分類することで得られる詳細な文書データ自体を分類した結果だけではなく、分割前の文書データと分類構造の組合せした結果とを基に、分割前の文書データを分類する分類手段と、前記分類構造を基に可能な文書分類手段が挙げられると、いう効果を奏する。

分類手段と、前記分割文分類手段による分類結果に基づいて分割文分類結果情報と生成する分割文分類結果生成手段と、前記文一分割文分類結果情報と前記分割文分類結果情報とを用いて前記文書データの分類結果情報と生成する文分類結果生成手段と、を備えるため、一つの文の中に複数の話題や意味が含まれており、一つの文の中に特定の話題や意味が含まれていない場合に、ある特定の話題や意味に関連する文がカテゴリー

分類されたり、利用者の意図するカテゴリとは異なるカテゴリに分類されたりすることがなく、したがって、利用者のその分類カテゴリをうまく理解が可能で、また、分割前文（所属文）中の分割文の位置なども示されるので、利用者は文群中の読みだした部分と効果的に隣むくことが可能な文分類装置が轉られるという効果を奏する。

【03993】また、請求項15の発明によれば、請求項14の発明において、前記文書データを保存する文書保存手段と、前記文書データを保存する文書保存手段と、前記文書データを保存する生成手段により生成された文一分割文 対応ワツプを保存する文書一分割文 対応ワツプを生成する手段と、分割文データおよび文 一分割文 対応ワツプを再生生成することにおいて、同一の文 データに対して、分類数、分類された要素付値情報とを用いて前記文書データを分類する構成に分類したため、文書データの詳細な分類構造がわかる文書分類を適用可能な文書分類装置が得られるという効果を奏する。

【03984】また、請求項20の発明によれば、請求項14〜17の発明において、前記文書分割手段が、指定された指定範囲にわたって文 データの分割をこなす構成に適用し、利用者の意図に合い、かつ文 データ

手法、または分類時の諸設定など、パラメータの異なる分類結果を効率的に求めることが可能で、また、文書テーマを分類し、分類結果を生成するために必要なデータが保持されることにより、利用者が分類作業に対して時間的な自由度を持つことが可能で、過去に行った文書分類の再分析を任意の時間間におこなうことも可能な文書分類装置が得られるという効果を奏する。

【03994】また、請求項16の発明によれば、請求項15の発明において、前記分割文書分類結果生成手段により生成された分割文書分類結果情報と保存された分割文書分類結果保存手段を備えるため、請求項15の発明の効果に加え、一度分類を実行すれば、その分類結果をテキスト表現や数表現やグラフ表現などさまざまな形式で表現することが可能で、また、分割文書分類結果情報が保存されることにより、分割の実行作業を自由なタイミングで再実行することが可能で、過去に行った文書分類結果を再分析することも可能で、さまざまな表現形式で任意の時間におこなうことも可能な文書分類装置が得られるという効果を奏する。

【0395】また、請求項17の発明によれば、請求項4～16の発明において、前記文書分類規則により生成される複数の分類文書データベースには、分割前の文書データベースのものを含め、利用者には、分割前の文書データベースを分類することで得られる詳細な文書データベースの分類構造だけでなく、分割前の文書データベース自体を分類した結果として得られる概略的でマクロな分類構造の組合せとした分類法を得ることが可能な文書分類装置と得られるという効果を奏する。

【0396】また、請求項18の発明によれば、請求項14～17の発明において、前記文書分析手段が、文書データの構造情報をもとに文書データを分割する構成のため、異なった話題の分割等を通知におこなうことができるが、異なつて、文書データの詳細な分類構造がわかれている文書分類を通知におこなうことが可能な文書分類装置が得られるという効果を奏する。

【03997】また、請求項19の発明によれば、請求項14～17の発明において、前記文書データに含まれる要素を抽出する文書要素抽出手段と、前記文書要素抽出手段により抽出された要素に付随する要素付随情報を出する要素付随情報抽出手段と、を備え、前記文書分割手段が、前記文書要素抽出手段により抽出された要素、または前記要素と前記要素付随情報抽出手段により抽出

それと、文書分類情報とを用いて前記文書データを分割する構成にしたため、文書データの詳細な分類結果がわかること、文書分類を適切におこなうことが可能な文書分類装置が得られるという効果を奏する。

【0398】また、請求項20の発明によれば、請求項14〜17の発明において、前記文書分割手段が、指示された指定範囲に基づいて、データデータの分類と、この構成にしたため、利用者の意図に合い、かつ文書データ

タの詳細な分類構造がわかる文書分類をおこなうことが可能な文書分類装置が得られるという効果を奏する。

【03093】また、請求項2の発明によれば、請求項14～17において、前記文書分類手段が、文書データ中の文字数、文数、または文字数と文数の両方を基に文書データを分類する構成にいたっては、話題の異なる文書データが異なる文書として分類される可能性が高くなるが、文書データの詳細な分類精度が得られるという効果を奏する。

【0400】また、請求項22の発明によれば、請求項14～21の発明において、前記文書分類結果生成手段が、文書データを示す情報および前記文書データに付随する代表的な情報、分類結果情報として抽出して提示する構成にしたため、利用者は文書データの詳細な分類構造的構成、分類結果の抽出結果を容易に把握することができるという効果を奏する。

【0400】また、請求項23の発明によれば、請求項22の発明において、前記文書分類装置は生成手段が、付随する文書データを示す情報および前記文書データについて、代表的情報を、分類結果情報として、抽出して提供する構成にしたため、利用者は文書データの詳細な分類構造の概要と全体的な構造とともにこの分類文書が従属している当該カテゴリに分類されたかというようなことも容易にわかる文書分類装置が得られるという効果を奏する。

【0402】また、請求項24の発明によれば、入力された文書データを記憶する文書記憶工程と、前記文書記憶工程により記憶された文書データの全部または一部を選択する選択工程と、前記選択工程により選択された文書データの全部または一部から文字列の特徴に関するデータを抽出する特徴抽出工程と、前記特徴抽出工程によ

り抽出された文字列の特徴に関するデータに基づいて前記文書データ、前記加工処理工程により加工処理された文書データと、前記加工処理工程により加工処理された文書データの全篇または一部を出力する出力工程と、をきめることにより、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文書処理方法が提供されるという効果を奏する。

〔40403〕 また、請求項25の発明によれば、前記出力工程が、前記加工処理工程により加工処理された文書データの全部または一部の内容に基づいて複数の項目値を設定する項目値設定工程と、前記項目値設定工程による設定された項目値ごとにも前記文書データの全部または一部を集計する集計工程と、を含み、前記文書データの全部または一部を、項目値を少なくとも二つの軸とする表形式に展開して出力するので、簡易な操作で加工処理の結果をクロス表として表すことができ、表の内容の把握を容易におこなうことができることから、文の意

味に添わるような分析作業において、単にその結果のみを出力するのではなく、情報分析作業全般にわたる支援をおこなうことが可能な文 処理方法が得られるという効果を奏する。

【10404】また、請求項26の発明に於ては、前記出力工程が、さらに、前記加工処理工程により加工処理された文書データの全部または一部を、前記加工処理工程により加工処理された前の文データの全部または一部とともに出力するので、加工処理すべし対象データとそれ以外のデータが同時に表示され、それを確認することにより、加工処理の対象範囲の決定を正確かつ容易におこなうことができて、文の意図に即応するようになつたことが、情報分析作業全般における支拂となつて、一般的な文書処理方法が得られるという効果とを奏する。

[0400] また、請求項27の発明によれば、前記文書加工機工程が、さらに、前記加工処理工程より加工処理された文書データの全部または一部を記憶するので、以後、他のデータと同様に扱うことができる。このため、文書の意味に係るような分析作業において、単にその結果のみを出力するのではなく、情報部分析作業全般にわたって文書データを扱うことが可能な文書処理方法が得られるという効果を奏する。

[0400] また、請求項26の発明によれば、前記型紙工程が、さらに、前記出力工程により出力された文データ全体の全部または一部を削除するので、出力工程により出力された文書データの全部または一部をさらに別の文書の対象とすることができ、多岐で高度な情報分析作業を行うことができることから、文の意味に係わるような分析作業において、単にその結果のみを出さるのではなく、作ることができること、また、その結果の出ることも可能な文書処理方法が得られるという効果を奏する。

【0407】また、請求項29の発明によれば、前記文書処理工程が、さらに、前記加工処理の内容に関するデータを記憶し、当該データの管理が容易になるようにデータの紛失を防止し、加工処理に用いた設定とそれによる処理結果を関連づけて記憶することができることから、文書の意味に気付くような分析作業において、単にその結果のみを出力することはなく、情報分析作業全体にわたる文書をおこなうことが可能な文書処理方法が得られるという効果を奏する。

〔0406〕また、請求項3000の発明によれば、入力工程が、文書データを入力し、直読解読工程が、前記入力工程より入力された文 データを解読して直読解読工程より生成された直読解読情報に基づいて分類工程 データに対する得られた直読解読情報に基づいて前記工程 データに対する文書資料工程より生成し、分類工程が、前記ベクトル生成工程より生成された文 データの特徴ベクトルと前記直読解読情報に基づいて文 を分類し、文 の部分集合を生成し、

クラス特徴算出工程が、前記分類工程により生成された文の部分集合の特徴であるクラス特徴を算出し、分類体系生成工程が、前記クラス特徴算出工程により算出されたクラス特徴に基づいて分類体系の構成要素を生成するので、クラスを得ることができるとともに、クラス重心間の類似度等を用いて、クラスの内容に基づくクラスの特徴化・体系化をおこなうことができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類方法が得られるという効果を奏する。

【0409】また、請求項31の発明によれば、入力工程が、文書データを入力し、言語解析工程が、前記入力工程により入力された文データを解析して言語解析情報を得、ベクトル生成工程が、前記言語解析工程により得られた言語解析情報に基づいて前記文書データに対する文特徴ベクトルを生成し、分類工程が、前記ベクトル生成工程により生成された文書特徴ベクトル間の類似度に基づいて文を分類し、文書の部分集合を生成し、クラス特徴算出工程が、前記分類工程により生成された文の部分集合の特徴であるクラス特徴を算出し、表示工程が、前記クラス特徴算出工程により算出されたクラス特徴を表示し、クラス選択指示工程が、前記分類工程により生成された文書部分集合の中から所望の部分集合を選択し、分類体系生成工程が、前記クラス選択指示工程により選択されたクラス特徴に基づいて分類体系の構成要素を生成するので、選択されたクラスのみを用いて、より操作者の意図したものに近いクラスを生成することができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類方法が得られるという効果を奏する。

【0410】また、請求項32の発明によれば、請求項31の発明において、ベクトル修正工程が、前記クラス選択指示手段により選択された部分集合に属する文書の文特徴ベクトルを除去したのりとなるように修正し、前記分類工程が、前記ベクトル修正工程により修正された文書特徴ベクトルに基づいて文書を分類するので、既知になったクラスの特徴を排除した新たなクラスを生成することができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類方法が得られるという効果を奏する。

【0411】また、請求項33の発明によれば、請求項31の発明において、文表現空間修正工程が、前記ベクトル生成工程により生成された文書特徴ベクトル間の類似度を判断する際、文表現空間を前記クラス選択指示工程により選択された部分集合から算出する特徴量に基づいて修正し、前記分類工程が、前記文表現空間修正工程により修正された文表現空間を用いて、前記ベクトル生成手段工程により生成された文特徴ベクトル間の類似度に基づいて文書を分類するので、前回の分類実行の結果、操作者に選択されたクラスの特徴を文の分類実行時に排除することができ、排除した状態で新たなクラスを生成することができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類方法が得られるという効果を奏する。

【0412】また、請求項34の発明によれば、請求項32の発明において、文書表現空間修正工程が、前記ベクトル生成工程により生成された文書特徴ベクトル間の類似度を判断する際、文表現空間を前記クラス選択指示工程により選択された部分集合から算出する特徴量に基づいて修正し、前記分類工程が、前記文書表現空間修正工程により修正された文書特徴ベクトル間の類似度に基づいて文書を分類するので、既知になったクラスの影響を排除し、かつ、前回の分類実行の結果、操作者に選択されたクラスの特徴を次の分類実行時に排除することができ、排除した状態で新たなクラスを生成することができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類方法が得られるという効果を奏する。

【0413】また、請求項35の発明によれば、請求項31または33の発明において、選択情報付与工程が、前記分類工程により生成された文書部分集合に所属する文書のすべてあるいは一部が選択された場合に選択されたことを示す選択情報を与え、前記表示工程が、前記クラス特徴を表示するとしても、選択情報付与工程により付与された選択情報を表示するので、多重に利用される文書の識別性および一度も選択されない文書の識別性を向上させることができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類方法が得られるという効果を奏する。

【0414】また、請求項36の発明によれば、請求項31～35の発明において、前記分類体系生成工程が、前記選択指示工程により選択されたクラス特徴の一部が、前記文書部分集合の中から選択された文書の部分集合に所属する文書群の全部あるいは一部および/または操作者が作成した情報に基づいて分類体系の構成要素を生成するので、クラスの内容や知識等を容易にし、かつ、操作者独自の分類体系を容易に生成することので、分類体系の利用面を向上させることができ、これにより、任意の文書集合にどのような内容が含まれるかを漸次的に収集することが可能な文書分類方法が得られるという効果を奏する。

【0415】また、請求項37の発明によれば、一つの文の中に複数の話題や意味が含まれている場合に、ある特定の話題や意味に関連されたカテゴリに分類された

り、利用者の意図するカテゴリとは異なるカテゴリに分類されたりすることがなく、したがって、利用者がその分類カテゴリをよく理解できる。また、分割前文書(所風文書)中の分割文書の位置なども示されるので、利用者は文書群中の読みたい部分を効率的に読むことが可能な文書分類方法が得られるという効果を奏する。

【0416】また、請求項38の発明によれば、請求項24～37のいずれか一つに記載された方法をコンピュータに実行させるプログラムを記録したことで、そのプログラムを機械読み取り可能となり、これによって、請求項24～37の動作をコンピュータによって実現することが可能な記録媒体が得られるという効果を奏する。

【図面の簡単な説明】
【図1】この発明の実施の形態1による文書処理装置を構成する情報処理システム全体のハードウェア構成を示す説明図である。
【図2】この発明の実施の形態1による文書処理装置を構成する情報処理システムにおけるサーバのハードウェア構成を示す説明図である。

【図3】この発明の実施の形態1による文書処理装置を構成する情報処理システムにおけるクライアントのハードウェア構成を示す説明図である。
【図4】この発明の実施の形態1による文書処理装置の構成を機能的に示すブロック図である。

【図5】この発明の実施の形態1による文書処理装置の項目名と項目値の関係を示す説明図である。
【図6】この発明の実施の形態1による文書処理装置の文書記憶部に記憶された文書のデータ構造を示す説明図である。

【図7】この発明の実施の形態1による文書処理装置の文書記憶部に記憶された文書の別のデータ構造を示す説明図である。
【図8】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。

【図9】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図10】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。

【図11】この発明の実施の形態1による文書処理装置の特徴抽出部によりおこなわれる抽出処理の内容の一覧を示す説明図である。
【図12】この発明の実施の形態1による文書処理装置の加工処理部によりおこなわれる加工処理の内容の一覧を示す説明図である。

【図13】この発明の実施の形態1による文書処理装置の各項目の特徴ベクトルを示す説明図である。
【図14】この発明の実施の形態1による文書処理装置の単語とその単語IDごとの出現回数を示す説明図である。

【図15】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図16】この発明の実施の形態1による文書処理装置の出力部によるクロス集計作成のための指示画面を示す説明図である。
【図17】この発明の実施の形態1による文書処理装置の出力部による分類処理の結果が表示されたクロス表を示す説明図である。

【図18】この発明の実施の形態1による文書処理装置の出力部による分類処理の結果が表示された別のクロス表を示す説明図である。
【図19】この発明の実施の形態1による文書処理装置の出力部の詳細な構成を示すブロック図である。

【図20】この発明の実施の形態1による文書処理装置のクロス表の出力手順を示すフローチャートである。
【図21】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図22】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。

【図23】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図24】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図25】この発明の実施の形態1による文書処理装置の文書記憶部の詳細な構成を示すブロック図である。

【図26】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図27】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図28】この発明の実施の形態1による文書処理装置の出力部による画面表示の別の例を示す説明図である。
【図29】この発明の実施の形態1による文書処理装置の文書処理の一連の手順を示すフローチャートである。

【図30】この発明の実施の形態2による文書処理装置の構成を機能的に示すブロック図である。
【図31】この発明の実施の形態2による文書処理装置のクラス特徴算出部の一列を示す説明図である。
【図32】この発明の実施の形態2による文書処理装置の一連の処理の手順を示すフローチャートである。
【図33】この発明の実施の形態3による文書処理装置の構成を機能的に示すブロック図である。

【図34】この発明の実施の形態3による文書処理装置の一連の処理の手順を示すフローチャートである。
【図35】この発明の実施の形態4による文書処理装置の構成を機能的に示すブロック図である。
【図36】この発明の実施の形態4による文書処理装置の一連の処理の手順を示すフローチャートである。
【図37】この発明の実施の形態5による文書処理装置の構成を機能的に示すブロック図である。

【図38】この発明の実施の形態5による文書処理装置の出力部による画面表示の別の例を示す説明図である。

一連の処理の手順を示すフローチャートである。

【図39】この発明の実施の形態6による文書分類装置の構成を概念的に示すブロック図である。

【図40】この発明の実施の形態6による文書分類装置の分類結果記憶部において取付けられたテーブルを示す説明図である。

【図41】この発明の実施の形態6による文書分類装置の選択情報付与部の処理の手順を示すフローチャートである。

【図42】この発明の実施の形態7を示す文書分類装置の構成ブロック図である。

【図43】この発明の実施の形態7による文書分類装置および文 分類方法の説明図である。

【図44】この発明の実施の形態7による文書分類装置および文 分類方法の他の説明図である。

【図45】この発明の実施の形態7による文書分類装置および文 分類方法の他の説明図である。

【図46】この発明の実施の形態7による文書分類装置および文 分類方法の他の説明図である。

【図47】この発明の実施の形態8による文書分類装置の構成ブロック図である。

【図48】この発明の実施の形態9による文書分類装置の構成ブロック図である。

【図49】この発明の実施の形態10による文書分類装置および文書分類方法の説明図である。

【図50】この発明の実施の形態11による文書分類装置および文 分類方法の説明図である。

【図51】この発明の実施の形態12による文書分類装置の構成ブロック図である。

【図52】この発明の実施の形態12による文書分類装置および文 分類方法の説明図である。

【図53】この発明の実施の形態13による文書分類装置および文 分類方法の説明図である。

【図54】この発明の実施の形態14による文書分類装置および文 分類方法の説明図である。

【図55】この発明の実施の形態15による文書分類装置および文 分類方法の説明図である。

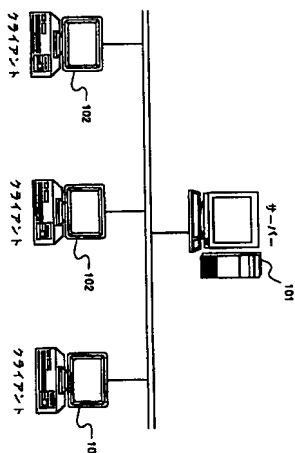
【図56】この発明の実施の形態16による文書分類装置および文 分類方法の説明図である。

【符号の説明】

- 101 サーバー
- 102 クライアント
- 103 ネットワーク
- 201 CPU
- 204 I/F
- 206 ディスク装置
- 301 CPU
- 306 ハードディスク
- 308 ディスプレイ
- 309 I/F

- 5006 文書分類結果生成部
- 5007 文書保存部
- 5008 分割文書保存部
- 5009 文書一分割文書対応なツツ保存部

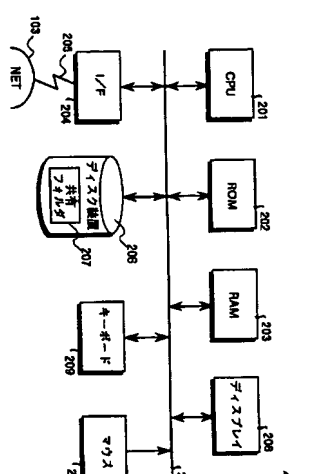
【図1】



【図12】

加工処理内容	
1	分類処理
2	検索処理
3	並び替え処理
4	代価算出処理
5	最大値算出処理
6	最小値算出処理
7	算術処理
:	:

【図2】



【図14】

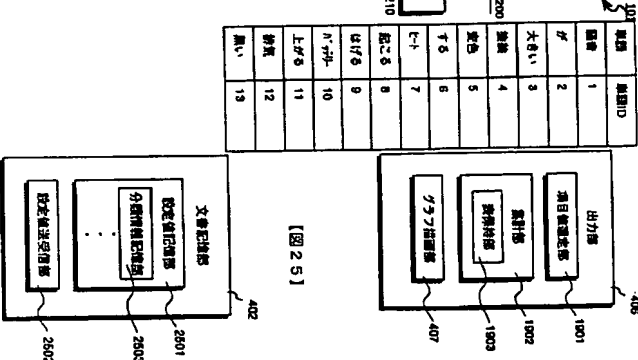
【図19】

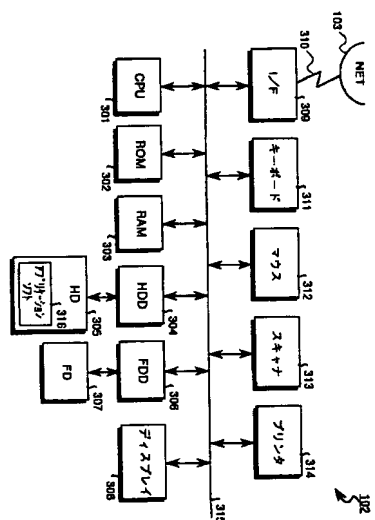
項目ID	項目名
1	大い
2	大い
3	大い
4	大い
5	大い
6	大い
7	大い
8	大い
9	大い
10	大い
11	大い
12	大い
13	大い

【図5】

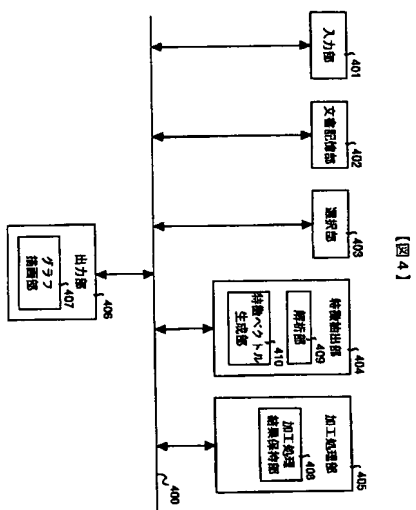
項目名	項目値
出版番号	特開平10-00000
出版日	平成10年1月1日
冊数の数	3
冊数の名称	本書の冊数
冊数の説明	本書の冊数は、(X)をXしているもので...

【図25】

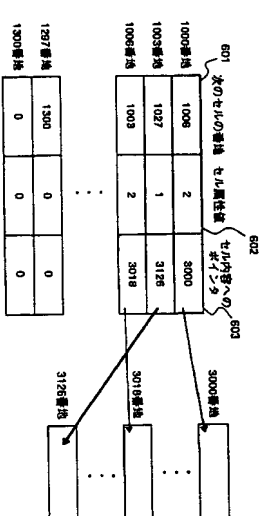




【例3】



【圖4】



【図6】

701 女子属性面 女子中身の 702 女子中身の 女子中身の 女子中身の

1000番地	2	3000	2	3018	2	3327
1100番地	1	3421	1	3512	1	3598

1500番地 0 0 0 0 0 0 0

【図7】

1500番地	0	0	0	0	...	0	0
--------	---	---	---	---	-----	---	---

【88】

番号	受付日	番付所	番種	年式	内容
801	1987/3/5	愛知	ABC1800	1983	所有者が大きい
1	1987/3/5	愛知	ABC1800	1985	所有者が大きい
2	1987/3/5	愛知	ABC1800	1985	所有者が大きい
3	1987/3/5	札幌	ABC1800	1986	所有者が大きい
4	1987/3/5	札幌	DEF1600	1986	所有者が大きい
5	1987/3/5	愛知	NLM1200	1982	所有者が大きい
6	1987/3/5	愛知	DEF1600	1984	所有者が大きい
7	1987/3/5	愛知	DEF1600	1986	所有者が大きい
8	1987/3/5	愛知	ABC1800	1987	所有者が大きい
9	1987/3/5	愛知	ABC1800	1982	所有者が大きい
10	1987/3/5	愛知	NLM1200	1982	所有者が大きい
11	1987/3/5	愛知	ABC1800	1984	所有者が大きい
12	1987/3/5	愛知	DEF1600	1986	所有者が大きい
13	1987/3/5	愛知	ABC1800	1987	所有者が大きい
14	1987/3/5	愛知	ABC1800	1985	所有者が大きい

＜クランク作成＞

知照対象車目名: 1601

知となる車目名: 1602

分限数: 50

1603 1604

【図6】

番号	受付日	製造所	車種	年式	内容
80	1987/7/8	大正	ABC1000	1984	押下げ悪い
81	1987/7/2	大正	ABC1800	1983	オキムシ食われる
82	1987/7/4	愛知	ABC1800	1983	エンジンが壊れる
83	1987/7/5	東京	ABC1800	1986	エンジンが壊れる
84	1987/7/6	東京	ABC1800	1987	エンジンが壊れる
85	1987/7/5	東京	ABC1800	1986	エンジンが壊れる
13	1987/7/10	大正	ABC1800	1984	エンジンが壊れる
28	1987/7/12	広島	ABC1800	1985	エンジンが壊れる
35	1987/7/17	大正	ABC1800	1985	エンジンが壊れる
38	1987/7/20	愛知	ABC1800	1985	エンジンが壊れる
42	1987/7/22	愛知	ABC1800	1985	エンジンが壊れる
46	1987/7/24	山梨	ABC1800	1984	エンジンが壊れる
47	1987/7/24	山梨	ABC1800	1984	エンジンが壊れる
48	1987/7/24	山梨	ABC1800	1985	エンジンが壊れる
49	1987/7/25	札幌	DEF1800	1985	エンジンが壊れる

【图 13】

【図10】

番号	受付日	番照所	車種	年式	内容
11	1987/3/6	大塚	ABC1800	1984	新車が大きい
53	1987/4/21	大塚	ABC1800	1985	新車が大きい
54	1987/4/21	大塚	ABC1800	1985	新車が大きい
1	1987/3/6	大塚	ABC1800	1984	新車が大きい
8	1987/3/6	大塚	ABC1800	1984	新車が大きい
13	1987/3/6	大塚	ABC1800	1984	新車が大きい
18	1987/3/10	大塚	ABC1800	1984	新車が大きい
28	1987/3/12	大塚	ABC1800	1984	新車が大きい
35	1987/3/17	大塚	ABC1800	1984	新車が大きい
39	1987/3/20	大塚	ABC1800	1984	新車が大きい
42	1987/3/22	大塚	ABC1800	1984	新車が大きい
46	1987/3/24	大塚	ABC1800	1984	新車が大きい
47	1987/3/24	大塚	ABC1800	1984	新車が大きい
4	1987/3/6	大塚	ABC1800	1984	新車が大きい

【図11】

1	対象とする文字列に含まれる車種
2	対象とする文字列に含まれる車種
3	対象とする文字列に含まれる車種の文字数
4	対象とする文字列に含まれる車種の出現回数
5	対象とする文字列に含まれる車種の出現回数
6	対象とする文字列に含まれる車種の出現回数
7	対象とする文字列に含まれる車種の出現回数
8	対象とする文字列に含まれる車種の出現回数
9	対象とする文字列に含まれる車種の出現回数
10	対象とする文字列に含まれる車種の出現回数

【図15】

番号	受付日	番照所	車種	年式	内容
1	1987/3/6	大塚	ABC1800	1984	新車が大きい
2	1987/3/6	大塚	ABC1800	1984	新車が大きい
3	1987/3/6	大塚	ABC1800	1984	新車が大きい
4	1987/3/6	大塚	ABC1800	1984	新車が大きい
5	1987/3/6	大塚	ABC1800	1984	新車が大きい
6	1987/3/6	大塚	ABC1800	1984	新車が大きい
7	1987/3/6	大塚	ABC1800	1984	新車が大きい
8	1987/3/6	大塚	ABC1800	1984	新車が大きい
9	1987/3/6	大塚	ABC1800	1984	新車が大きい
10	1987/3/6	大塚	ABC1800	1984	新車が大きい
11	1987/3/6	大塚	ABC1800	1984	新車が大きい
12	1987/3/6	大塚	ABC1800	1984	新車が大きい
13	1987/3/6	大塚	ABC1800	1984	新車が大きい
14	1987/3/6	大塚	ABC1800	1984	新車が大きい

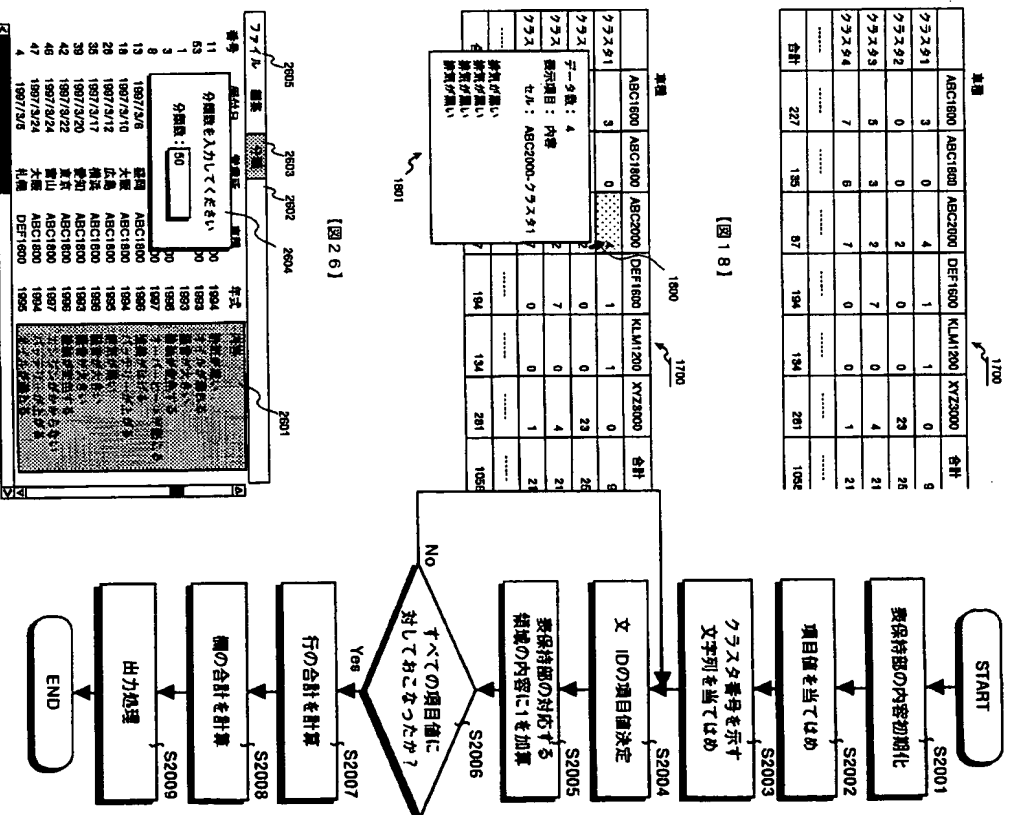
【図17】

車種	ABC1800	ABC1800	ABC2000	DEF1600	KLM1200	XYZ2000	合計
クラス91	3	0	4	1	1	0	9
クラス92	0	0	2	0	0	23	25
クラス93	5	3	2	7	0	4	21
クラス94	7	6	7	0	0	1	21
合計	227	185	87	154	154	281	1002

【図18】

車種	ABC1800	ABC1800	ABC2000	DEF1600	KLM1200	XYZ2000	合計
クラス91	3	0	4	1	1	0	9
クラス92	0	0	2	0	0	23	25
クラス93	5	3	2	7	0	4	21
クラス94	7	6	7	0	0	1	21
合計	227	185	87	154	154	281	1002

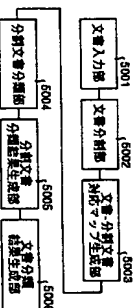
【図20】



【図26】

番号	受付日	番照所	車種	年式	内容
11	1987/3/6	大塚	ABC1800	1984	新車が大きい
53	1987/4/21	大塚	ABC1800	1985	新車が大きい
54	1987/4/21	大塚	ABC1800	1985	新車が大きい
1	1987/3/6	大塚	ABC1800	1984	新車が大きい
8	1987/3/6	大塚	ABC1800	1984	新車が大きい
13	1987/3/6	大塚	ABC1800	1984	新車が大きい
18	1987/3/10	大塚	ABC1800	1984	新車が大きい
28	1987/3/12	大塚	ABC1800	1984	新車が大きい
35	1987/3/17	大塚	ABC1800	1984	新車が大きい
39	1987/3/20	大塚	ABC1800	1984	新車が大きい
42	1987/3/22	大塚	ABC1800	1984	新車が大きい
46	1987/3/24	大塚	ABC1800	1984	新車が大きい
47	1987/3/24	大塚	ABC1800	1984	新車が大きい
4	1987/3/6	大塚	ABC1800	1984	新車が大きい

【図22】



【図21】

番号	受付日	登録所	重畳	形式	内容
1	1997/3/5	愛知	ABC1800	1993	図面が大きい
2	1997/3/5	愛知	ABC2000	1995	図面が大きい
3	1997/3/5	東京	ABC1800	1996	図面が大きい
4	1997/3/5	札幌	DEF1800	1996	図面が大きい
5	1997/3/5	福岡	KLM1200	1992	図面が大きい
6	1997/3/5	愛知	DEF1800	1994	図面が大きい
7	1997/3/5	愛知	DEF1800	1996	図面が大きい
8	1997/3/5	愛知	ABC1800	1992	図面が大きい
9	1997/3/5	愛知	XYZ2000	1997	図面が大きい
10	1997/3/5	愛知	KLM1200	1992	図面が大きい
11	1997/3/5	愛知	DEF1800	1997	図面が大きい
12	1997/3/5	愛知	DEF1800	1997	図面が大きい
13	1997/3/5	愛知	ABC1800	1996	図面が大きい
14	1997/3/5	愛知	XYZ2000	1996	図面が大きい

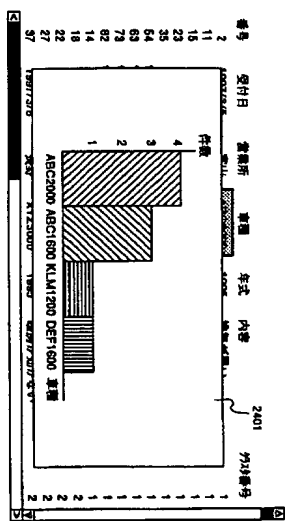
【図22】

番号	受付日	登録所	重畳	形式	内容
1	1997/3/5	愛知	ABC1800	1993	図面が大きい
2	1997/3/5	愛知	ABC2000	1995	図面が大きい
3	1997/3/5	東京	ABC1800	1996	図面が大きい
4	1997/3/5	札幌	DEF1800	1996	図面が大きい
5	1997/3/5	福岡	KLM1200	1992	図面が大きい
6	1997/3/5	愛知	DEF1800	1994	図面が大きい
7	1997/3/5	愛知	DEF1800	1996	図面が大きい
8	1997/3/5	愛知	ABC1800	1992	図面が大きい
9	1997/3/5	愛知	XYZ2000	1997	図面が大きい
10	1997/3/5	愛知	KLM1200	1992	図面が大きい
11	1997/3/5	愛知	DEF1800	1997	図面が大きい
12	1997/3/5	愛知	DEF1800	1997	図面が大きい
13	1997/3/5	愛知	ABC1800	1996	図面が大きい
14	1997/3/5	愛知	XYZ2000	1996	図面が大きい

【図23】

番号	受付日	登録所	重畳	形式	内容
1	1997/3/5	愛知	ABC1800	1993	図面が大きい
2	1997/3/5	愛知	ABC2000	1995	図面が大きい
3	1997/3/5	東京	ABC1800	1996	図面が大きい
4	1997/3/5	札幌	DEF1800	1996	図面が大きい
5	1997/3/5	福岡	KLM1200	1992	図面が大きい
6	1997/3/5	愛知	DEF1800	1994	図面が大きい
7	1997/3/5	愛知	DEF1800	1996	図面が大きい
8	1997/3/5	愛知	ABC1800	1992	図面が大きい
9	1997/3/5	愛知	XYZ2000	1997	図面が大きい
10	1997/3/5	愛知	KLM1200	1992	図面が大きい
11	1997/3/5	愛知	DEF1800	1997	図面が大きい
12	1997/3/5	愛知	DEF1800	1997	図面が大きい
13	1997/3/5	愛知	ABC1800	1996	図面が大きい
14	1997/3/5	愛知	XYZ2000	1996	図面が大きい

【図24】



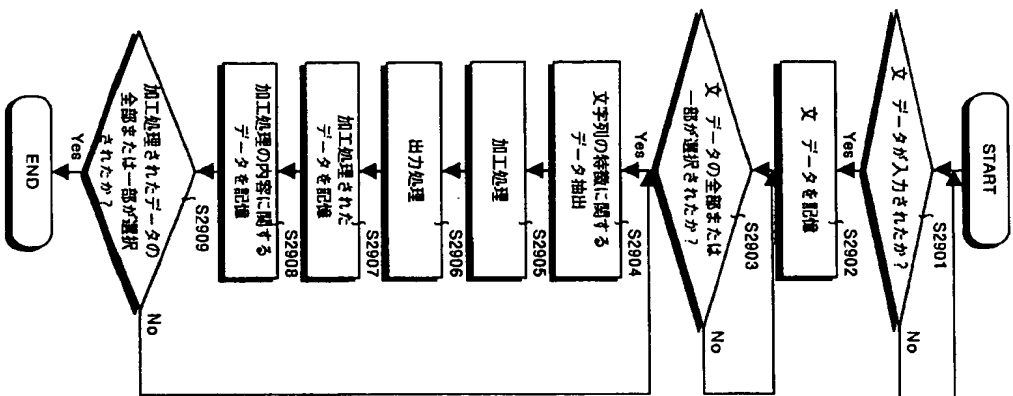
【図27】

番号	受付日	登録所	重畳	形式	内容
1	1997/3/5	愛知	ABC1	1997/5/12日	図面が大きい
2	1997/3/5	愛知	ABC2	1997/5/12日	図面が大きい
3	1997/3/5	愛知	ABC1	1997/5/12日	図面が大きい
4	1997/3/5	愛知	ABC2	1997/5/12日	図面が大きい
5	1997/3/5	愛知	ABC1	1997/5/12日	図面が大きい
6	1997/3/5	愛知	ABC2	1997/5/12日	図面が大きい
7	1997/3/5	愛知	ABC1	1997/5/12日	図面が大きい
8	1997/3/5	愛知	ABC2	1997/5/12日	図面が大きい
9	1997/3/5	愛知	ABC1	1997/5/12日	図面が大きい
10	1997/3/5	愛知	ABC2	1997/5/12日	図面が大きい
11	1997/3/5	愛知	ABC1	1997/5/12日	図面が大きい
12	1997/3/5	愛知	ABC2	1997/5/12日	図面が大きい
13	1997/3/5	愛知	ABC1	1997/5/12日	図面が大きい
14	1997/3/5	愛知	ABC2	1997/5/12日	図面が大きい

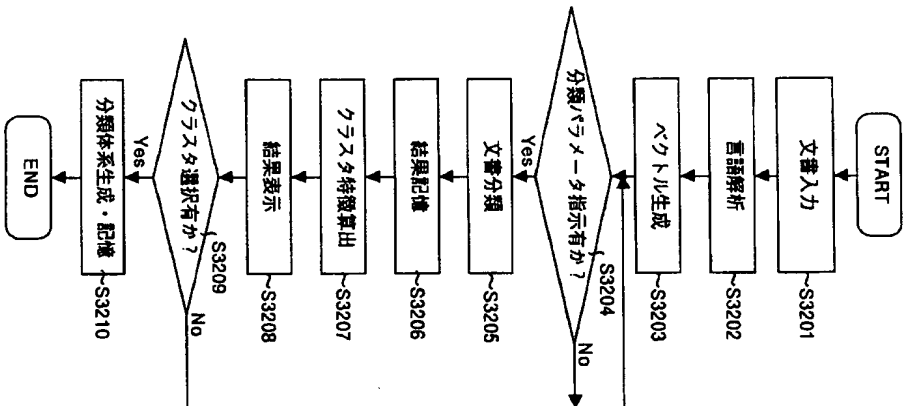
【図28】

番号	受付日	登録所	重畳	形式	内容
1	1997/3/5	愛知	ABC1800	1993	図面が大きい
2	1997/3/5	愛知	ABC2000	1995	図面が大きい
3	1997/3/5	東京	ABC1800	1996	図面が大きい
4	1997/3/5	札幌	DEF1800	1996	図面が大きい
5	1997/3/5	福岡	KLM1200	1992	図面が大きい
6	1997/3/5	愛知	DEF1800	1994	図面が大きい
7	1997/3/5	愛知	DEF1800	1996	図面が大きい
8	1997/3/5	愛知	ABC1800	1992	図面が大きい
9	1997/3/5	愛知	XYZ2000	1997	図面が大きい
10	1997/3/5	愛知	KLM1200	1992	図面が大きい
11	1997/3/5	愛知	DEF1800	1997	図面が大きい
12	1997/3/5	愛知	DEF1800	1997	図面が大きい
13	1997/3/5	愛知	ABC1800	1996	図面が大きい
14	1997/3/5	愛知	XYZ2000	1996	図面が大きい

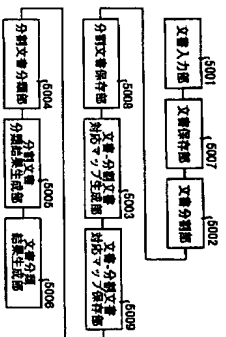
【図 2 9】



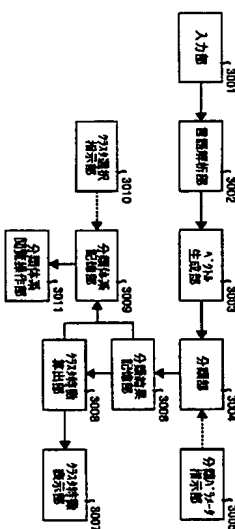
【図 3 2】



【図 4 7】



【図 3 0】

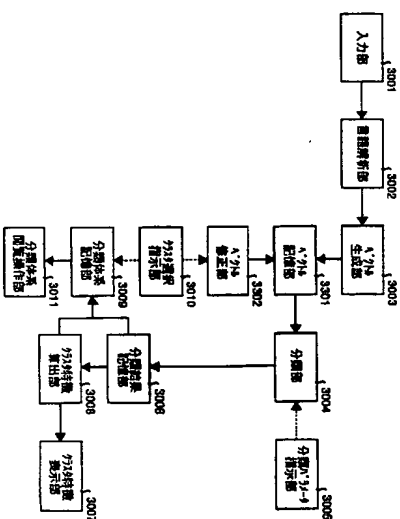


【図 3 1】

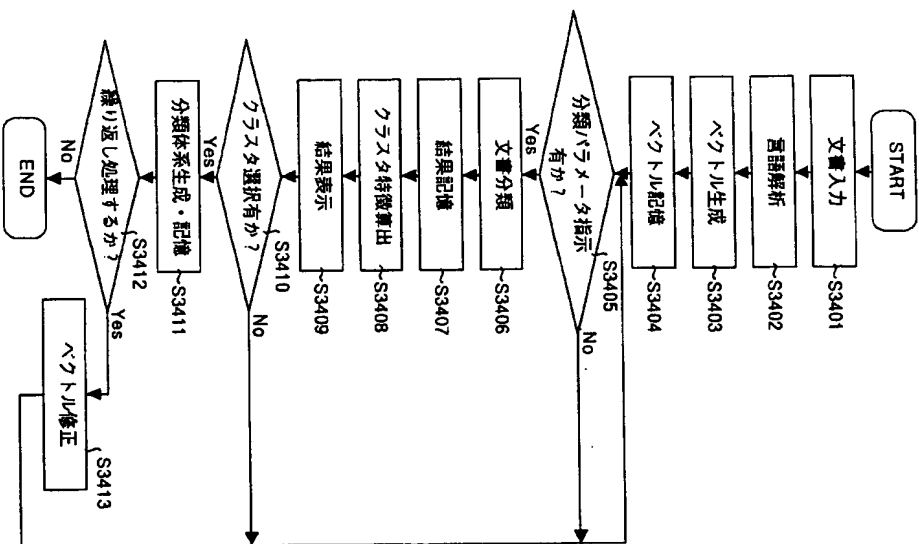
3101	3102	3103	3104	3105
1	248	管理費、多代...	文書内容	文書内容
2	249	管理費、多代...	文書内容	文書内容
3	250	管理費、多代...	文書内容	文書内容
4	251	管理費、多代...	文書内容	文書内容
5	252	管理費、多代...	文書内容	文書内容
6	253	管理費、多代...	文書内容	文書内容
7	254	管理費、多代...	文書内容	文書内容
8	255	管理費、多代...	文書内容	文書内容
9	256	管理費、多代...	文書内容	文書内容
10	257	管理費、多代...	文書内容	文書内容

N	1408	操作性、悪い・・・	777の操作性が悪い	0.989
			777の操作性を算える大変である	0.882

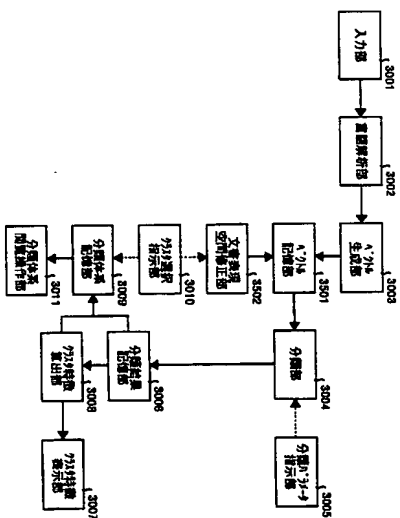
【図 3 3】



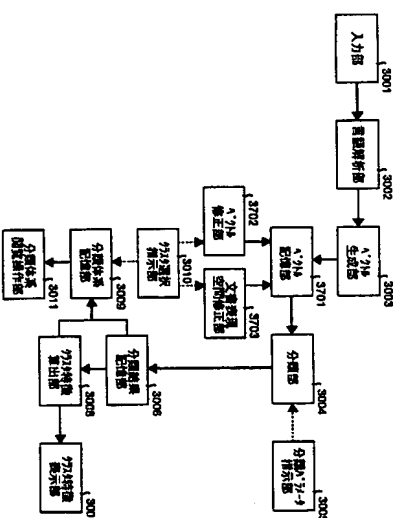
【図 3 4】



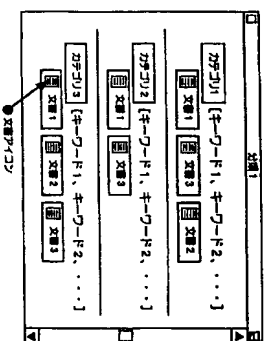
【図 3 5】



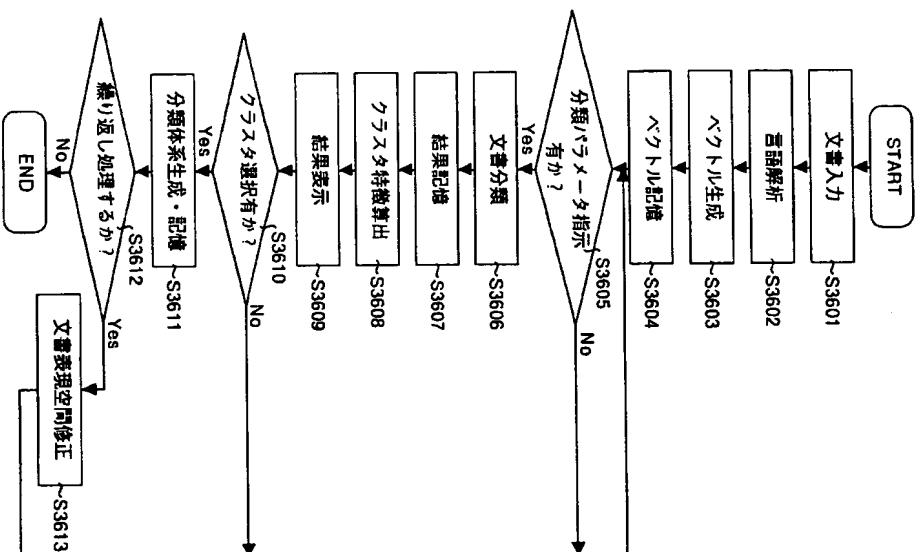
【図 3 7】



【図 5 5】



【図36】



【図37】

